UNIVERSIDAD PERUANA DE CIENCIAS APLICADAS

FACULTAD DE CIENCIAS DE LA SALUD

PROGRAMA ACADÉMICO DE MEDICINA

Modelos predictivos de muerte o exacerbación en Insuficiencia Cardiaca

Aguda: Una revisión sistemática.

TESIS

Para optar el título profesional de Médico cirujano

AUTOR

Rojas Ortega, Alex (0000-0002-5013-1792)

Arias Reyes, Francisco Alejandro (0000-0003-3805-9166)

ASESOR

Hernández Díaz, Adrián Vladimir (0000-0002-9999-4003)

Steyerberg, Ewout Willem (0000-0002-7787-0122)

Benites Zapata, Vicente Aleixandre (0000-0002-9158-1108)

Lima. 13 de abril del 2020

DEDICATORIA

Dedicamos este trabajo a nuestras familias, que siempre nos apoyaron a lo largo de los años para poder cumplir nuestros sueños. A nuestros profesores que siempre nos recalcaron nuestra importante labor y la empatía que nos permita entender como se sienten los demás. Y nuestros amigos, que a lo largo de estos años se volvieron nuestra segunda familia y nuestra mayor fuente de información sobre como afrontar la vida.

AGRADECIMIENTOS

Agradecemos a nuestros asesores, que nos guiaron a lo largo del presente trabajo. Sin su ayuda no hubiera sido posible concluirlo.

RESUMEN

1. Introducción y Objetivos: Existe un gran número de modelos predictivos de desenlaces en Insuficiencia Cardíaca Aguda; sin embargo, no se conoce la utilidad clínica. Se evaluará el desempeño de los modelos incluidos en el artículo.

2. Metodología: Se realizó una revisión sistemática de los modelos predictivos de desenlaces adversos. Por lo que se realizó una búsqueda en 6 bases bibliográficas, incluyéndose todo artículo original que reporte modelos de predicción. Se utilizó la herramienta PROBAST para valorar el riesgo de sesgo en los artículos.

3. Resultados: De 2498 artículos seleccionados, 39 fueron incluidos. La mayoría provienen de Europa (46.1%), 20 de ellos evalúan mortalidad, mientras que 21 demostraron bajo riesgo de sesgo y buena aplicabilidad utilizando la herramienta de evaluación PROBAST. El sexo masculino fue el mayoritario en el 69% de artículos. La morbilidad más común es hipertensión arterial. Todos los modelos fueron validados, y más del 50% cuentan con validación externa. Borovac JA et al. presenta el AUC más alto, con 0.907. Catorce modelos utilizaron el modelo de calibración de Hosmer-Lemeshow. Ningún artículo contó con beneficio neto.

4. Conclusión:

: Los estudios no muestran una adecuada evaluación de desempeño, al no presentar todas las estadísticas de medición que existen; es necesario el desarrollo de nuevos modelos con la descripción completa de estadísticas de desempeño, para llevar a cabo una adecuada comparación.

5. Palabras Clave: Modelo Predictivo, Insuficiencia Cardíaca Aguda, Desempeño, Aplicabilidad.

ABSTRACT

1. Introduction and Objectives: There are currently a number of predictive models of outcomes in Acute Heart Failure (AHF); the clinical usefulness and the possible implementation of these are unknown. In this study, we propose to evaluate the performance of the predictive models for AHF.

2. Methodology: A systematic review of the predictive models of adverse outcomes was performed. Therefore, a search was made in 6 bibliographic databases, including all original articles that report prediction models. The PROBAST tool was used to assess the risk of bias in the articles.3. Results: From 2498 initial articles, we took 39 from them. The majority came from Europe (46.1%), 20 of them assess mortality, while 21 demonstrated low risk of bias and good applicability using the PROBAST assessment tool. The male sex was the majority in 69% of articles. The most common morbidity is high blood pressure. All the models were validated, and more than 50% have external validation. Borovac JA et al. it has the highest AUC with 0.907. Fourteen models used the Hosmer Lemeshow calibration model. No article had a net benefit.

4. Conclusion: The studies don't show an adequate evaluation of performance, since they do not present all the existing measurement statistics; The development of new models with the complete description of performance statistics is necessary to carry out an adequate comparison.5. Key Words: Predictive Model, Acute Heart Failure, Performance, Applicability.

Tabla de Contenidos

TABLA DE CONTENIDOS	6
ÍNDICE DE TABLAS.....	7
ÍNDICE DE FIGURAS.....	8
CAPÍTULO I: INTRODUCCIÓN.....	9
CAPÍTULO II: DESCRIPCIÓN DEL ESTUDIO	12
METODOLOGÍA:	12
<i>Tipo de estudio y criterios de inclusión.....</i>	<i>12</i>
<i>Proceso de Selección.</i>	<i>12</i>
<i>Identificación e información de variables principales</i>	<i>13</i>
<i>Descripción de los procedimientos del estudio.....</i>	<i>14</i>
<i>Aspectos Éticos.....</i>	<i>15</i>
CAPÍTULO III: RESULTADOS	16
CAPÍTULO IV: DISCUSIÓN	18
CAPÍTULO V: REFERENCIAS	21
BIBLIOGRAFÍA:.....	21
ANEXOS	72

Índice de tablas

Tabla 1: Características poblacionales de los modelos predictivos incluidos en la revisión sistemática.....	30
Tabla 2: Características de diseño de los modelos predictivos incluidos en la revisión sistemática.....	47
Tabla 3: Estadísticas de desempeño de los modelos predictivos incluidos en la revisión sistemática.....	55
Tabla 4: Resultados de evaluación de riesgo de sesgo utilizando la herramienta PROBAST.	¡Error! Marcador no definido.

Índice de Figuras

Figura 1: Flujograma de artículos seleccionados	29
---	----

CAPÍTULO I: INTRODUCCIÓN

La insuficiencia cardiaca (IC) es un síndrome heterogéneo que se define como la falla o descompensación circulatoria en la que el corazón no puede responder y satisfacer las necesidades del organismo. Esta patología causa una morbilidad importante en la población general, ya que 1 de cada 5 pacientes que desarrolla un Infarto Agudo de Miocardio (IMA) se complica con IC dentro del 1er mes por lo cual son readmitidos¹. Uno de los tipos de IC es la insuficiencia cardiaca aguda (ICA), una entidad que aparece luego del IMA, shock, o arritmias^{2,3}, o como la exacerbación de una patología crónica no controlada. La demora en el diagnóstico o manejo terapéutico de la ICA puede ser mortal. No se conoce prevalencia de la ICA posiblemente por la dificultad de definir una población con las características necesarias dentro de los artículos^{1,4,5,6}.

En efecto, en nuestro país se han escrito trabajos sobre Insuficiencia Cardiaca Aguda, aunque la información es muy escasa. Se ha estudiado características epidemiológicas y clínicas en pacientes con ICA.⁷ Se encontró que el sexo femenino es el mayoritario entre estos pacientes, además de HTA e IMA como principales comorbilidades. Un tercio de ellos ya había presentado ICA previamente. El artículo de Pariona, et al. nos da información con respecto a la ICA en un hospital de alta complejidad, presentando que el 16% de los pacientes hospitalizados en el servicio de Cardiología tienen el diagnóstico de ICA al ingreso.⁸ En dicho artículo se encontró que la HTA e IMA son las principales causas de ICA; y que en este grupo de pacientes la mortalidad intrahospitalaria que ascendió al 7%. En otro artículo realizado en el interior del país, han probado utilizar la puntuación de CHADS2, utilizada en fibrilación auricular, para evaluar su desempeño al aplicarla en pacientes con ICA cuyo desenlace sea la muerte; sin embargo, se encontró que este no predecía adecuadamente la mortalidad.⁹

En la actualidad, existen distintos modelos predictivos de ICA, los cuales utilizan diversas variables para estimar el riesgo de muerte en pacientes con esta patología^{10,11}. Los modelos predictivos de ICA muestran variables como el historial médico (ej. edad, raza), comorbilidades, etiología, síntomas clásicos de la ICA, signos vitales, hallazgos físicos a la exploración y pruebas de apoyo al diagnóstico (ej. Pro-PNC, Radiografía de tórax)¹². El artículo de Fonarow GC, et al de 2008⁶ muestra dos modelos predictivos: el EFFECT risk tool y el ADHERE risk tool. Estos modelos han demostrado utilidad diagnóstica y estratificación del riesgo de mortalidad, pero presentan ciertas dificultades, como no tener una población donde validar los modelos; o que los modelos no presentan indicadores de desempeño (performance). La herramienta EFFECT⁶ consta de 22 variables, y busca estratificar a pacientes con riesgo bajo y alto de mortalidad al mes y al año: quienes tengan un puntaje bajo (<60) tienen 0,4% de muerte en un mes, y 7,8% en un año; aquellos con puntaje alto (>150) tienen un porcentaje de casi el

60% de muerte en un mes y casi del 80% en un año. La herramienta ADHERE⁶ consta de 3 variables: BUN (Blood Urea Nitrogen por sus siglas en inglés), creatinina y presión arterial sistólica; y clasifica a los pacientes con bajo, intermedio y alto riesgo de mortalidad en la emergencia⁵. Existen otros modelos predictivos para ICA tales como las herramienta PROTECT¹³ y STRATIFY^{1,14}; estos modelos difieren de los previos en que evalúan otras variables, como la muerte, reanimación cardiopulmonar, diálisis, entre otros; y su relación con otra variable no estudiada previamente: la troponina. Los desenlaces más estudiados en estos artículos son la mortalidad en un plazo de tiempo: 1 semana, 30 días y 1 año^{13,14}.

Existen diferencias al momento de evaluar modelos predictivos de ICA en comparación a modelos predictivos de insuficiencia cardiaca crónica (ICC) debido a que los modelos predictivos de ICC toman valores de cronicidad de enfermedad. En los modelos predictivos de ICC se toma en cuenta otros parámetros tales como fracción de filtración glomerular, niveles de colesterol (LDL, HDL) y consumo crónico de medicamentos. Estos parámetros hacen la diferencia entre modelos predictivos de insuficiencia cardiaca aguda (ICA) y crónica (ICC)^{15,16}. En este punto, también es importante resaltar que la cantidad de pacientes con ICA para estudiar comparado con la población de ICC es menor. Los datos son más complejos de recolectar en pacientes con sospecha de ICA que llegan a una emergencia por un problema agudo como exacerbación de una valvulopatía o IMA, y con riesgo de muerte¹⁷.

Existe actualmente literatura respecto a revisiones sistemáticas de los modelos predictivos de desenlaces adversos de insuficiencia cardiaca aguda; sin embargo, a diferencia del presente estudio se lleva la búsqueda sistemática en menos bases de datos, se toman en cuenta menos desenlaces, no se utiliza la herramienta PROBAST, específica para revisiones sistemáticas de modelos predictivos, no mencionan el beneficio neto como estadística importante desempeño, o no diferencia entre las entidades de un paciente con ICA e ICC.^{18, 19 y 20}

Como hemos mencionado previamente, la mayoría de los modelos predictivos estudiados evalúan la presión arterial sistólica, sexo o edad; que, ajustadas con otras variables como el BUN, están asociadas con mortalidad. Además, todos los modelos evalúan riesgo de mortalidad en siete días, un mes y/o año. Sin embargo, hay aspectos controversiales, dado que en algunos modelos pronósticos estudiados no se registra correctamente el desenlace de muerte de los pacientes, o incluso no mencionan cómo se manejan los datos perdidos de pacientes; por lo que es una información perdida; además de que el impacto de los modelos falta ser demostrado. También, todos los artículos de modelos predictivos buscan desarrollar y validar el modelo en los mismos pacientes, mas no aplicarlo a una población diferente^{1,21,22,23}.

Aunque existe un alto número de modelos predictivos de ICA, no se conoce si estos pueden llegar a una aplicabilidad en el ámbito hospitalario. Por tal motivo, el objetivo del trabajo es hacer una revisión exhaustiva de la literatura científica a fin de conocer el desempeño y aplicabilidad clínica de los

diferentes modelos predictivos de desenlaces como muerte y exacerbación en insuficiencia cardíaca aguda existentes alrededor del mundo.

CAPÍTULO II: Descripción del Estudio

Metodología:

Tipo de estudio y criterios de inclusión.

Diseño de estudio

Revisión sistemática siguiendo la guía PRISMA (en inglés, Preferred Reporting Items for Systematic Reviews and Meta-Analyses)^{24,25}.

Criterios de Inclusión

Se incluyó todo artículo científico en cualquier idioma que reporte uno o más modelos de predicción utilizando un análisis multivariable para evaluar el desarrollo y al menos uno de los desenlaces (Muerte, Exacerbación de síntomas de ICA, Exacerbación intrahospitalaria y reingreso hospitalario) de una insuficiencia cardiaca aguda en población adulta. Además, debe contar con al menos una estadística de evaluación de desempeño, reporte de la discriminación, calibración, reclasificación y/o utilidad clínica (descrita como beneficio neto o aplicabilidad clínica), y algún tipo de validación de los modelos desarrollados^{26,27}.

Criterios de exclusión

Se excluyó artículos editoriales, resúmenes, cartas al editor, revisiones sistemáticas y estudios de datos duplicados. Tampoco se tuvieron en cuenta aquellos artículos que trataban sobre modelos de predicción de insuficiencia cardiaca, sin especificar que fuera la entidad aguda.

Proceso de Selección.

Tiempo de Estudio

La búsqueda y selección de artículos se dio entre los meses de enero, febrero y marzo del 2018, además de una actualización de la búsqueda en agosto del 2019.

Estrategia de Búsqueda

Se realizó la búsqueda estratégica de artículos relacionados al tema de investigación en 6 buscadores (PubMed, Web of Science, EMBASE, Cochrane library y Scopus), con la estrategia de búsqueda avanzada descrito en el Anexo1; no se aplicó ninguna

restricción respecto a los idiomas de los modelos que participarán del estudio. Posterior a ello se revisó de forma independiente el título y resumen de los artículos obtenidos por dos de los autores (ARO, FAR) de manera independiente, quienes a su vez compararon los resultados obtenidos para discutir las discrepancias con ayuda del tercer autor (AVH). Toda la lista de artículos seleccionados por títulos y resúmenes fue descargada a texto completo a fin de aplicarles los criterios de elegibilidad y quedarse con los artículos que han sido incluidos en la revisión para la extracción de datos y análisis.

Identificación e información de variables principales

Variable de exposición

Variables que forman parte del modelo multivariado, el cual es la base del score clínico o modelo predictivo. Evaluamos la estrategia de selección de variables y número de variables ^{26,27}.

Variable de resultado

Desenlaces evaluados en modelos de predicción ^{1,13}.

Muerte: Cese de los signos vitales del paciente en un plazo de 7, 30 días o 1 año posteriores al cuadro de insuficiencia cardíaca aguda.

Reingreso Hospitalario: Re-entrada al hospital en un plazo de 7, 30 días o 1 año posteriores al cuadro de insuficiencia cardíaca aguda.

Exacerbación de síntomas de Insuficiencia cardíaca: Aumento transitorio de gravedad de síntomas o aparición de nuevos, posteriores al evento de insuficiencia cardíaca aguda en un paciente con antecedente de insuficiencia cardíaca crónica¹¹.

Exacerbación Intra-Hospitalaria: Progresión hacia un estado de gravedad en donde es necesario una de las siguientes condiciones: Aplicación de inotrópicos o un vasodilatador intravenoso >12 horas después de la presentación en el hospital; transferencia a la unidad de cuidados intensivos, utilización de atención terapéutica médica avanzada (es decir, soporte circulatorio mecánico, ventilación mecánica o hemodiálisis) después del primer día de ingreso²⁸.

Estadísticas de desempeño en los modelos predictivos de ICA²⁵

Discriminación: Capacidad de un modelo de predicción para distinguir entre los pacientes que desarrollan y no desarrollan el resultado.

Calibración: Exactitud de un modelo de predecir las probabilidades de riesgos, e indica en qué medida están de acuerdo los resultados esperados (previstos por el modelo) y los resultados observados²⁹.

Reclasificación: Porcentaje de la población del estudio que es reclasificada en un diferente desenlace ante una variación en el número de predictores²⁹.

Índice de reclasificación neta (NRI): Cuantifica el número de pacientes que están correctamente reclasificados al usar el nuevo marcador frente al uso del marcador tradicional.

Índice de mejora de discriminación integrada (IDI): Representa lo que mejora en promedio el nuevo modelo en cuanto a la predicción de más verdaderos eventos, descontando lo que empeora por la predicción de falsos eventos.

Utilidad Clínica - Beneficio Neto: Relación entre el beneficio y riesgo de la aplicación de un modelo pronóstico en la práctica clínica²⁷.

Descripción de los procedimientos del estudio:

Se extrajeron de manera independiente las variables previamente descritas de los estudios seleccionados por dos autores de manera independiente (ARO, FAR). Las discrepancias fueron resueltas con la ayuda del tercer autor (AVH). Estas variables coinciden con las que existen en la lista de verificación para la evaluación crítica y extracción de datos para revisiones sistemáticas de modelos predictivos o CHARMS por sus siglas en inglés^{19,22,23}. Los elementos que se extrajeron de los artículos que describen el desarrollo del modelo incluyeron: el diseño del estudio (Cohorte, ensayo clínico), población de estudio, ubicación geográfica, resultado, horizonte de predicción, método de modelado (modelo de riesgos proporcionales de Cox, modelo logístico, lineal), y medidas de rendimiento predictivo (calibración, discriminación, reclasificación, beneficio neto), número de participantes en el estudio, número y tipo de predictores, y tipo de desenlace^{25,29,30}.

Se hizo la distribución de los datos en tablas utilizando Microsoft Excel 2016 donde se registraron las variables principales: región del mundo, tamaño de población, edad media, % de mujeres y comorbilidades de insuficiencia cardíaca. Para la presentación de los datos de discriminación se usaron las estadísticas de concordancia o c-statistics, calibración a través del Hosmer-lemeshow, O: E, Beneficio Neto y reclasificación con IDI y NRI. El tipo de validación fue considerado como interno y externo dependiendo de la información brindada por el artículo primario.

Análisis del riesgo de sesgo

Para la evaluación dentro de los estudios extraídos se utilizó la Herramienta PROBAST.^{31,32} Esta comprende el uso de 4 pasos para reconocer el riesgo de sesgo y de aplicabilidad de los modelos pronósticos incluidos. Inicialmente busca establecer la pregunta de investigación, luego continua con el tipo de modelo incluido y 4 dominios que constan de 20 preguntas donde se evalúa si fue correcta la selección de participantes, desenlace, predictores y análisis; finalmente el último paso comprende la evaluación final donde se concluye si el modelo presentado cuenta con riesgo de sesgo o si es aplicable, por lo cual se clasificó de acuerdo a lo sugerido por Wolff, et al. Si el modelo presentaba alto riesgo de sesgo en al menos uno de los dominios, se le consideraba en la evaluación general como estudio con alto riesgo de sesgo, mientras que si se encontraba un dominio con datos poco exactos o imprecisos, aún si todos los otros fueran bajos; se consideraba a este como estudio con sesgo indeterminado. Además, si un estudio presentaba todos los dominios bajos pero este no tenía ninguna clase de evaluación, se le consideraba como estudio de alto riesgo de sesgo^{31,32}.

Aspectos Éticos

El trabajo fue evaluado por el comité de ética de la Universidad Peruana de Ciencias Aplicadas, siendo exonerado al tratarse de una revisión sistemática con el expediente PI223-17.

Existen beneficios con respecto a este trabajo, y es que nos ayudará a tener en cuenta el posible mejor modelo predictivo de muerte o exacerbación en Insuficiencia Cardíaca Aguda.

CAPÍTULO III: Resultados

Se obtuvieron 2498 artículos seleccionados posterior a la búsqueda sistemática, respecto al número por cada base de datos: 585 a través de PubMed, 513 en EMBASE, 244 Cochrane Library, 372 en Scopus y 784 en Web of Science. De los cuales 412 fueron eliminados por tratarse de duplicados. 2086 pasaron a revisión de resúmenes; de los cuales 1994 no se incluyeron por no tener concordancia con la pregunta de investigación de la revisión, al no tratarse de desarrollos ni validaciones de modelos predictivos. De los 95 restantes, 53 artículos fueron eliminados, 39 por no brindar o poder acceder a información completa sobre datos generales de población o del desarrollo del modelo, 12 por ser artículos descriptivos sobre modelos predictivos, dos por tratarse revisiones sistemáticas. El número de artículos incluidos en la revisión fue de 39 (Figura 1).

Las características de los estudios seleccionados se encuentran en la tabla 1. Encontramos que en orden de frecuencia, demográficamente, la mayor producción de modelos predictivos es validada en población del continente europeo (46.1%) seguido por Asia (30.7%), EE. UU (15.3%), Canadá (2.5%); dos artículos no dieron datos demográficos. Respecto al tamaño de población, el artículo con el mayor número fue el desarrollado por De Vore A, et al con 15640, mientras que el de menor tamaño fue el de Lourenco P, et al con 224 participantes. El tiempo de seguimiento mínimo desarrollado para un modelo fue de 7 días, mientras que el mayor fue de 5 años. En cuanto a la edad el estudio que presento la media más alta fue el de Arenja N, et al con 82 años promedio; mientras que el más bajo fue de 62 años promedio en los estudios de Panerica J, et al y Scrutinio D, et al. Además la mayoría de los modelos (69%) fueron desarrollados o validados en una población en su mayoría de sexo masculino. La morbilidad que más estuvo presente fue la hipertensión arterial, encontrándose en más del 50% de la población en casi la totalidad de artículos incluidos.

La tabla 2 extrae ciertas características de diseño de los modelos predictivos seleccionados, y denota que la mayoría de trabajos son cohortes retrospectivas y prospectivas. La mayoría toma como variables de exposición a la PA (mas comúnmente la presión arterial sistólica), sexo, edad, algunos hallazgos laboratoriales tales como sodio en sangre, enzimas cardíacas, niveles de creatinina y ácido úrico, consumo previo de medicamentos y comorbilidades, entre otras. Las estrategias de selección, en la población de artículos, es de asociaciones encontradas en algunas variables, o encontradas en artículos previos que fueron citados. Algunos toman variables de estudios previos como el de Chen YJ et al., quien utiliza el modelo AHEAD y lo relaciona con ácido úrico. Formiga F et al., Jabov J et al. Kinugasa Y et al, Lee DS et al y Hsieh M et al. no presentan estrategias de selección. Se utiliza comúnmente la

Regresión de Cox y regresiones logísticas, como también algunos modelos no presentan los métodos de modelo, como Gil V et al. y Hsieh M et al. Finalmente, los modelos suelen ser de desarrollo aunque también hay un grupo importante que genera también validaciones. Todos tienen validez y más del 50% de modelos hacen también validación externa.

La tabla 3 presenta las estadísticas de desempeño de los modelos predictivos incluidos en la revisión sistemática. Respecto a los desenlaces encontramos 20 modelos que evaluaban mortalidad independientemente del lugar, 6 con evaluación de mortalidad o reingreso hospitalario, 6 con evaluación de mortalidad intrahospitalaria, 2 con evaluación Mortalidad o exacerbación de síntomas, 2 con evaluación de mortalidad, exacerbación o reingreso hospitalario, 2 con evaluación de solo reingreso intrahospitalario y 1 con evaluación de exacerbación intrahospitalaria. La mayoría de artículos incluidos cuentan con al menos un método de evaluación de desempeño, siendo el de mayor prevalencia la discriminación a través de las curvas ROC; siendo el de mejor desempeño el desarrollado por Borovac JA, et al³³ con AUC : 0.907(CI95% :0.867- 0.939), mientras que los de menor desempeño fueron los de DeVore A, et al²⁸ en su validación externa: 0.63 y Leong k, et al³⁴ en su modelo básico: 0.647(CI95% :0.62-0.674). Solo 14 de los modelos seleccionados presentaron datos de calibración a través de Hosmer-Lemeshow, de estos la validación externa del modelo de DeVore A, et al²⁸ fue el único que rechazó la hipótesis nula, es decir que el modelo no tenía adecuada bondad de ajuste. Se presenta la reclasificación en 9 modelos incluidos, siendo el NT-ProBNP el biomarcador que más se adicionó, mostrando mejores resultados. Ninguno de los modelos presentó el beneficio neto dentro de sus métodos de análisis de desempeño.

La tabla 4 muestra la valoración de riesgo de sesgo de los estudios incluidos en la revisión. Del total de modelos (que contaron 40 al final, debido a que se realizaron 2 riesgos de sesgos con la herramienta PROBAST en el modelo de Chen et al, por tener 2 outcomes distintos), 21 cuentan con un grado de bajo riesgo de sesgo, además de buena aplicabilidad y 2 cuentan con un grado bajo de riesgo, pero con una baja evaluación de aplicabilidad. Por el contrario 11 modelos presentaron alto riesgo de sesgo, pero buena aplicabilidad, 2 presentaron alto riesgo y una baja evaluación de aplicabilidad 4 presentaron datos insuficientes para la evaluación de riesgo de sesgo, pero contaban con una adecuada aplicabilidad y por último 1 presentó datos insuficientes tanto para la evaluación de riesgo de sesgo y aplicabilidad. A su vez, los datos poco certeros o inexistentes en la presentación de los artículos seleccionados, mostraron un alto índice de sesgo a nivel del dominio de análisis, evaluado con la herramienta PROBAST^{31,32}.

CAPÍTULO IV: Discusión

En la presente investigación se evidencia que la mayoría de los estudios incluidos hablan sobre el desarrollo de un nuevo modelo predictivo en comparación a los que buscan la validación externa de uno ya existente; así mismo son pocos los que incluyen ambas características al mismo tiempo. También podemos observar que según su frecuencia demográfica la mayoría de ellos fueron realizados en Europa, seguido por Asia y EE. UU, siendo su aplicación sobre la raza negra o mestiza casi nula. Más de la mitad de ellos no muestran datos sobre el desempeño como la calibración y discriminación; además ninguno muestra la utilización de Beneficio neto ²⁷ lo que nos daría una mejor idea de su utilidad. Dentro de los modelos más estudiados contamos con el estudio ADHERE^{6,34}, EFFECT^{35,36} y GWTG-HF^{37,38}, los cuales han servido de base para la creación de nuevos modelos, así como validaciones externas. Respecto a las variables predictoras, los modelos incluidos tienen un número que usualmente oscila entre seis a nueve, existiendo algunas excepciones en donde existen solo tres variables y teniendo como máximo 21 variables. La variable edad estuvo presente en la mayoría de los modelos. En frecuencia le sigue los signos cuantificables en el paciente o presentes como antecedentes en la historia clínica como la presión arterial sistólica y las comorbilidades. El resto de los componentes suelen ser exámenes de laboratorio, siendo los más recurrentes el examen de nitrógeno urémico en la sangre, la prueba de péptidos natriuréticos y la creatinina sérica; y en menor medida pruebas de imágenes como los ecocardiogramas.

Respecto a publicaciones previas relacionadas a la ICA y modelos predictivos, hemos podido encontrar estudios de comparación desarrollados por Lagu T, et al³⁹ y Chirinos D, et al⁴⁰, los cuales incluyen 7 y 2 modelos respectivamente; sin embargo, a diferencia del presente trabajo, no cuentan con una búsqueda sistemática ni evaluación de riesgo de sesgo, por lo cual no podemos estar seguros del nivel de evidencia de los modelos⁶. En la presente revisión hemos considerado todos los modelos predictivos de ICA en todas las bases de datos, además de añadir todos los desenlaces posibles. A su vez, se ha utilizado el PROBAST^{31,32}, herramienta creada para evaluar el riesgo de sesgo en revisiones sistemáticas sobre modelos predictivos; esto difiere de la revisión de Michaud AM, et al⁴¹, donde sólo se utiliza a la mortalidad como desenlace y utiliza la escala Newcastle-Ottawa (NOS por su nombre en inglés) para medir el riesgo de sesgo.

Las falencias de la mayoría de los modelos presentados es que carecen de validación externa o tienen más de 10 años publicados sin actualización o nueva validación. Además, se han intentado juntar modelos no específicos para ICA o buscan modelos que tengan desenlace de tiempo igual para compararlos^{39,40,41}.

La revisión de la aplicabilidad se encuentra relacionada a las características y número de variables predictoras seleccionadas por estudio. Dentro de todos ellos, debemos resaltar el NT-Pro-BNP score^{42,43}, en el cual se utiliza el NT-Pro-BNP y el cambio de sus valores a lo largo de la estancia hospitalaria como predictores de mortalidad en los pacientes; debido a que reporta uno de los valores más altos en las curvas⁴³. Además presentó una adecuada bondad de ajuste (Hosmer-Lemeshow); esto difiere de la mayoría de modelos en los cuales se utilizan múltiples variables y no lograron obtener valores aceptables de bondad de ajuste^{31,32,37,40}. Dentro de los modelos estudiados con el mejor desempeño de calibración tenemos al reportado en el estudio de Yagyu T, et al⁴⁴. Este toma como referencia el modelo desarrollado por Peterson P, et al⁴⁵, el cual cuenta con 7 variables predictoras de recolección fácil,⁴⁵ además de añadirle los cambios dinámicos de valores a lo largo del tiempo a lo largo estancia intrahospitalaria del paciente, lo que permite predecir no solo el pronóstico a la llegada a la emergencia sino qué se puede esperar a lo largo del manejo; esto difiere de la mayoría de estudios incluidos, los cuales suelen presentar una evaluación estática sin una evaluación del estado del paciente posterior al manejo inicial.

La utilidad de esta revisión es amplia para los médicos; en especial los cardiólogos. El Infarto Agudo de Miocardio (IAM) es una de las principales causas de ICA, y ésta es muy común en nuestro país y el mundo¹. Se sabe que es una de las primeras causas de muerte en la población en general; y es relevante saber o conocer cuál es el estado de los pacientes que llegan a las emergencias; cuál es su riesgo muerte en la atención inicial y cuál es su riesgo de muerte tiempo después de salir de alta. La utilización de estos modelos debería ser imprescindible en todos los pacientes con ICA, ya que es vital estratificar el riesgo, y adoptar así las medidas de cuidado correspondientes, dependiendo del riesgo de mortalidad. Este uso debería darse tanto en la emergencia como en ambiente intrahospitalario. La revisión que hemos realizado nos ayuda a conocer cuál modelo deberíamos seguir, en medida de conseguir uno con pocas variables, variables rápidas y fáciles de obtener, principalmente en el ámbito de una emergencia en un hospital, pero también en pacientes ya hospitalizados. Por ello, el modelo de Yagyu Y, et al⁴⁴, el cual cuenta con 7 variables sencillas y rápidas de obtener, además de ser un modelo validado-interna y externamente, debería ser el modelo preferido por los cardiólogos, emergencistas; y médicos en general que evalúen pacientes con ICA.

El impacto en salud pública de estos modelos podría ser positivo, en la medida que estos se adopten. Las medidas que se tomen en los pacientes que presenten riesgos de muerte, sean altos o bajos, en los días siguientes al diagnóstico de ICA serán más rápidas y exactas, además de tener un plan de seguimiento de los pacientes antes de salir de alta maximizando la mejor calidad en salud de los pacientes.

Una de las más importantes limitaciones fue la heterogeneidad marcada en los estudios incluidos en la revisión, lo que no permite realizar un meta-análisis y entonces poder obtener una medida de resumen de desempeño de los modelos evaluados. Esto probablemente se deba a las diferencias en la utilización de las variables en los modelos pronósticos y también a que muchos de los modelos incluidos muestran pocos datos respecto al desempeño y validación inconclusa. Además, se suma la falta de estudios cardiovasculares en diferentes poblaciones; la mayoría de los estudios se llevan a cabo en la raza caucásica^{47,48,49,50,51}, con validaciones externas en cohortes de Asia; sin embargo, no hemos encontrado trabajos que repliquen los modelos en poblaciones de Sudamérica o África; lo que no nos permite tener una representación clara del desempeño en mayor espectro de razas. Otra limitación es que la mitad de los artículos incluidos no cumplen con todos los requisitos que pide el PROBAST. Diecisiete de ellos cumplen criterios de riesgo alto, lo que dificulta el nivel de evidencia que pueden brindar.

En conclusión, sabemos que existen muchos modelos predictivos en insuficiencia cardiaca aguda que han sido publicados a lo largo de los años, pero son poco utilizados en nuestro medio. En el ámbito local, las condiciones en las emergencias no son las mejores y muchas veces no se cuenta con pruebas diagnósticas como ecocardiografía o pruebas laboratoriales como NT-proBNP. Lo que se necesita son modelos que prescindan de variables complejas, sean cortas y rápidas, sin requerir materiales costosos. Yagyu Y, et al⁴⁴ utiliza un modelo predictivo de muerte donde aplica 7 variables rápidas y no muy complejas de obtener en la emergencia. Está puede ser nuevamente evaluada durante la evolución hospitalaria; así mismo, el presente modelo cuenta con validación interna y externa en una población extensa, lo que nos permite valorar el riesgo de muerte de estos pacientes de forma más confiable.

Recomendamos que en una futura investigación sobre modelos predictivos se realicen modelos que tengan menos variables laboratoriales y más clínicas, que se busque un patrón predictivo entre variables que tengan una validación externa para asegurar la validez del modelo en una población grande, y que pueda extrapolarse en más poblaciones.

Para finalizar, sugerimos que en los distintos hospitales de nuestro país se desarrollen y/o apliquen modelos predictivos de muerte u otro desenlace en Insuficiencia Cardiaca Aguda, sabiendo pues que esta patología es muy común en nuestra población.⁸

CAPÍTULO V: Referencias

Bibliografía:

1. Collins SP, Jenkins CA, Harrell FE, Liu D, Miller KF, Lindsell CJ, et al. Identification of Emergency Department Patients With Acute Heart Failure at Low Risk for 30-Day Adverse Events: The STRATIFY Decision Tool. *JACC Heart Fail.* 2015; 3(10): 737–747.
2. Ponikowski P, Jankowska EA. Pathogenesis and Clinical Presentation of Acute Heart Failure. *Rev Esp Cardiol.* 2015;68: 331-7.
3. Weintraub N, Collins S, Pang P, et al. Acute heart failure syndromes. *Circulation* 2010;122:1975-96.
4. Nieminen M, Brutsaert D, Dickstein K, Drexler H, Follath F, Harjola VP, et al. EuroHeart Failure Survey II (EHFS II): a survey on hospitalized acute heart failure patients: description of population, *European Heart Journal.* 2006; 27(22): 2725–2736,
5. Searle J, Frick J, Möckel M. Acute heart failure facts and numbers: acute heart failure populations. *ESC Heart Failure* 2016; 3: 65–70
6. Fonarow GC. Epidemiology and risk stratification in acute heart failure. *Am Heart J* 2008; 155:200-207.
7. Michaud AM, Parker SIA, Ganshorn H, Ezekowitz JA, McRae AD. Prediction of Early Adverse Events in Emergency Department Patients With Acute Heart Failure: A Systematic Review. *Can J Cardiol.* 2018;34(2):168-179.
8. Rahimi K, Bennett D, Conrad N, Williams TM, Basu J, Dwight J, et al. Risk prediction in patients with heart failure: a systematic review and analysis. *JACC Heart Fail.* 2014;2(5):440-6.
9. Mahajan SM, Heidenreich P, Abbott B, Newton A, Ward D. Predictive models for identifying risk of readmission after index hospitalization for heart failure: A systematic review. *Eur J Cardiovasc Nurs.* 2018 Dec;17(8):675-689.

10. Loiacono F, Fragasso G, Calori G, Alberti L, Marinosci G, Salerno A, et al. Validation of a new score for outcome prediction in patients with heart failure with reduced ejection fraction. *Minerva Cardioangiol.* 2019;67(3):191-199
11. Rossello X, Miró O, Llorens P, Jacob J, Herrero-Puente P, Gil V, et al. Effect of Barthel Index on the Risk of Thirty-Day Mortality in Patients With Acute Heart Failure Attending the Emergency Department: A Cohort Study of Nine Thousand Ninety-Eight Patients From the Epidemiology of Acute Heart Failure in Emergency Departments Registry. *Annals of Emergency Medicine.* 2019;73(6):589 - 598
12. O'Connor CM, Mentz RJ, Cotter G, Metra M, Cleland JG, Davison BA, et al. The PROTECT in-hospital risk model: 7-day outcome in patients hospitalized with acute heart failure and renal dysfunction. *European Journal of Heart Failure.* 2012;14:605–612.
13. Lee D, Alba A. Risks and Benefits of Risk Prediction in Acute Heart Failure. *JACC: Heart failure.* 2015;3(10):748 – 50.
14. Passantino A, Monitillo F, Iacoviello M, Scrutinio D. Predicting mortality in patients with acute heart failure: Role of risk scores. *World Journal of Cardiology* 2015; 7(12): 902–911.
15. Aceña A, Martín-Mariscal M, Cristóbal C, et al. Comparison of 3 Predictive Clinical Risk Scores. *Texas Heart Institute Journal.* 2017;44(4).
16. Lee D, Austin P, Rouleau J, Liu P, Naimark D, Tu J. Predicting Mortality Among Patients Hospitalized for Heart Failure. Derivation and Validation of a Clinical Model. *JAMA.* 2003;290: 2581-2587.
17. Sakhnini A, Saliba W, Schwartz N, Bisharat N. The derivation and validation of a simple model for predicting in-hospital mortality of acutely admitted patients to internal medicine wards. *Medicine.* 2017; 96(25): e7284.16.
18. Villar M. Características epidemiológicas y clínicas de la Insuficiencia Cardíaca Aguda en un Hospital de Referencia, Arequipa, Perú 2017-2018. Universidad Católica de Santa María. Facultad de Medicina Humana. 2019
19. Pariona M, Segura Saldaña PA, Padilla Reyes M, Reyes Villanes JS, Jáuregui Contreras M, Valenzuela-Rodríguez G. Características clínico epidemiológicas de la insuficiencia cardíaca aguda en un hospital terciario de Lima, Perú. *Rev Peru Med Exp Salud Publica.* 2017;34(4):655-59. doi: 10.17843/rpmesp.2017.344.2890

20. Gonzáles C. Mendoza C. Evaluacion de puntaje predictivo CHADS2 para mortalidad en Insuficiencia Cardiaca Hospital Nacional Ramiro Prialé Prialé. Essalud Huancayo. 2017. Universidad Nacional del Centro del Perú. Facultad de Medicina Humana. 2018.
21. Dickstein K et al. ESC guidelines for the diagnosis and treatment of acute and chronic heart failure 2008: the Task Force for the diagnosis and treatment of acute and chronic heart failure 2008 of the European Society of Cardiology. Developed in collaboration with the Heart Failure Association of the ESC (HFA) and endorsed by the European Society of Intensive Care Medicine (ESICM). *Eur J Heart Fail*. 2008 Oct;10(10):933-89.
22. Ferrero P , Iacovoni A, D'Elia E, Vaduganathan M, Gavazzi A, Senni M. Prognostic scores in heart failure - Critical appraisal and practical use. *International Journal of Cardiology*. 2015 ;188:1-9.
23. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JP, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. *BMJ*. 2009;339:b270. Disponible en: <https://doi.org/10.1136/bmj.b2700>
24. Debray TP , Damen JA , Snell KI, Ensor J, Hooft L, Reitsma JB, et al. A guide to systematic review and meta-analysis of prediction model performance. *BMJ*. 2017; 356 :i6460.
25. Berhe W. Sahle, Alice J. Owen, Ken Lee Chin, Christopher M. Reid. Risk Prediction Models for Incident Heart Failure: A Systematic Review of Methodology and Model Performance. *Journal of Cardiac Failure*. 2017. 23(9):680-687.
26. Gonzales A. Selección de variables: Una revisión de métodos existentes. Universidade da Coruña. Master en técnicas estadísticas. 2015. 87p.
27. Vickers AJ, Calster BV, Steyerberg EW. Net benefit approaches to the evaluation of prediction models, molecular markers, and diagnostic tests. *BMJ*. 2016;352:i6.
28. DeVore AD, Greiner MA, Sharma PP, Qualls LG, Schulte PJ, Cooper LB, et al. Development and validation of a risk model for in-hospital worsening heart failure from the Acute Decompensated Heart Failure National Registry (ADHERE). *Am Heart J*. 2016; 178:198-205.
29. 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.

30. Alba AC, Agoritsas T, Jankowski M, Courvoisier D, Walter SD, Guyatt GH. Risk Prediction Models for Mortality in Ambulatory Patients With Heart Failure A Systematic Review *Circulation: Heart Failure*. 2013;6:881-889.
31. Wolff R, Whiting P, Mallett S, et al. PROBAST: a risk of bias tool for prediction modelling studies. *Cochrane Colloquium Vienna*. 2015.
32. Wolff RF, Moons KG, Riley RD, Whiting PF, Westwood M, Collins GS, et al. PROBAST: A Tool to Assess the Risk of Bias and Applicability of Prediction Model Studies. *Annals of Internal Medicine*. 2019;170(1):51-60.
33. Borovac JA, Glavas D, Bozic J, Novak K. Predicting the 1-Year All-Cause Mortality After Hospitalization for an Acute Heart Failure Event: A Real-World Derivation Cohort for the Development of the S2PLiT-UG Score. *Heart, Lung and Circulation*. 2019; S1443-9506(19)30351-8. doi: 10.1016/j.hlc.2019.03.021. Disponible en: [https://www.heartlungcirc.org/article/S1443-9506\(19\)30351-8/fulltext](https://www.heartlungcirc.org/article/S1443-9506(19)30351-8/fulltext)
34. Lim N-K, Lee SE, Lee H-Y, Cho H-J, Choe W-S, Kim H, et al. Risk prediction for 30-day heart failure-specific readmission or death after discharge: Data from the Korean Acute Heart Failure (KorAHF) registry. *Journal of Cardiology*. 2019;73:108–113.
35. Diercks DB, Fonarow GC, Kirk D, Emerman CL, Hollander JE, Weber JE, et al. Risk Stratification in Women Enrolled in the Acute Decompensated Heart Failure National Registry Emergency Module (ADHERE-EM). *ACADEMIC EMERGENCY MEDICINE* 2008; 15:151–158.
36. Thorvaldsen T, Claggett BL, Shah A, Cheng S, Agarwal SK, Wruck LM, et al. Predicting Risk in Patients Hospitalized for Acute Decompensated Heart Failure and Preserved Ejection Fraction The Atherosclerosis Risk in Communities Study Heart Failure Community Surveillance. *Circ Heart Fail*. 2017;10:e003992.
37. Spinar J, Jarkovsky J, Spinarova L, Mebazaa A, Gayat E, Vitovec J, et al. AHEAD score — Long-term risk classification in acute heart failure. *International Journal of Cardiology*. 2016;202:21–26.
38. Chen YJ, Sung SH, Cheng HM, Huang WM, Wu CL, Huang CJ, et al. Performance of AHEAD Score in an Asian Cohort of Acute Heart Failure With Either Preserved or Reduced Left Ventricular Systolic Function. *J Am Heart Assoc*. 2017

39. Lagu T, Pekow PS, Shieh MS, et al. Validation and Comparison of Seven Mortality Prediction Models for Hospitalized Patients With Acute Decompensated Heart Failure. *Circ Heart Fail*. 2016;9(8):10.1161/CIRCHEARTFAILURE.115.002912 e002912. doi:10.1161/CIRCHEARTFAILURE.115.002912
40. Chirinos D, Rinaldi M, Aguayo R, Trejo G, Leonardi M, Guthmann M, et al. Usefulness of Different Mortality risk Scores in Patients with Heart Failure. A retrospective observational Study. *REV ARGENT CARDIOL* 2018;86:322-328. <http://dx.doi.org/10.7775/rac.v86.i5.13212>
41. Michaud AM, Parker SIA, Ganshorn H, Ezekowitz JA, McRae AD, Prediction of early adverse events in emergency department patients with acute heart failure: A systematic review, *Canadian Journal of Cardiology* (2017), doi: 10.1016/j.cjca.2017.09.004.
42. Pérez-Calvo JI, Ruiz-Ruiz FJ, Carrasco Sanchez FJ, Morales-Rull JL, Manzano-Fernández S, et al Prognostic value of serum cystatin C and N-terminal pro b-type natriuretic peptide in patients with acute heart failure. *European Journal of Internal Medicine*.2012, 23(7):599 – 603
43. Eurlings LW, Sanders-Van Wijk S, Van Kraaij DJ, Van Kimmenade R, Meeder JG, Kamp O, et al. Risk Stratification With the Use of Serial N-Terminal Pro-B-Type Natriuretic Peptide Measurements During Admission and Early After Discharge in Heart Failure Patients: Post Hoc Analysis of the PRIMA Study. *Journal of Cardiac Failure*. 2014; 20(12): 881-890.
44. Yagyu T, Kumada M, Nakagawa T. Novel risk stratification with time course assessment of in-hospital mortality in patients with acute heart failure. *PLoS ONE*. 2017;12(11): e0187410. <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0187410>
45. Peterson PN, Rumsfeld JS, Liang L, Albert NM, Hernandez AF, et al. A Validated Risk Score for In-Hospital Mortality in Patients With Heart Failure From the American Heart Association Get With the Guidelines Program. 2010.3(1)
46. Arenja N, Breidhardt T, Socrates T, Schindler C, Heinisch C, Tschung C, et al. Risk stratification for 1-year mortality in acute heart failure. *Swiss Med Wkly*. 2011;141:w13259. Disponible en: <https://smw.ch/article/doi/smw.2011.13259>
47. Behnes M, Bertsch T, Weiss C, Ahmad-Nejadd P, Akin I, Fastner C, et al. Triple head-to-head comparison of fibrotic biomarkers galectin-3, osteopontin and gremlin-1 for long-term prognosis in suspected and proven acute heart failure patients. *International Journal of Cardiology*. 2016; 203:398–406.

48. Boralkar KA, Kobayashi Y, Moneghetti KJ, Pargaonkar VS, Tuzovic M, Krishnan G, et al. Improving risk stratification in heart failure with preserved ejection fraction by combining two validated risk scores. *Open Heart* 2019;6: e000961. Disponible en: <https://openheart.bmj.com/content/6/1/e000961>
49. Cameli M, Pastore MC, De Carli G, Henein MY, Mandoli GE, Lisi E, et al. ACUTE HF score, a multiparametric prognostic tool for acute heart failure: A real-life study. 2019; 296: 103- 108.
50. Delgado A, Rodrigues B, Nunes S, Baptista R, Marmelo B, Moreira D, et al. Acute Heart Failure Registry: Risk Assessment Model in Decompensated Heart Failure. *Arq Bras Cardiol.* 2016; 107(6):557-567
51. Formiga F, Masip J, Chivite D and X Corbella. Applicability of the heart failure Readmission Risk score: A first European study. *International Journal of Cardiology.* 2017; 236:304–309.
52. Garcia-Gutierrez JS, Quintana M, Antón-Ladislao A, Gallardo M, Pulido E, Rilo I, et al. Creation and validation of the acute heart failure risk score: AHFRS. *Intern Emerg Med.* 2017; 12:1197–1206.
53. Gil V, Miró O, Schull MJ, Llorens P, Herrero-Puente P, Jacob J, et al. Emergency Heart Failure Mortality Risk Grade score performance for 7-day mortality prediction in patients with heart failure attended at the emergency department: validation in a Spanish cohort. *European Journal of Emergency Medicine.* 2018; 25:169–177.
54. Hsieh M, Auble TE, Yealy DM. Validation of the Acute Heart Failure Index. *Ann Emerg Med.* 2008;51:37-44.
55. Jacob J, Miró Ò, Herrero P, Martín-Sánchez FJ, Gil V, Toste J, et al. Predicting short-term mortality in patients with acute exacerbation of chronic heart failure: The EAHFE-3D scale. *Med Intensiva.* 2016;40(6):348-355.
56. Jia Q, Wang YR, He P, Huang XL, Yan W, Mu Y, et al. Prediction model of in-hospital mortality in elderly patients with acute heart failure based on retrospective study. *Journal of Geriatric Cardiology.* 2017; 14: 669-678.
57. Khanam SS, Choi E, Son J-W, Lee J-W, Youn YJ, Yoon J, et al. Validation of the MAGGIC (Meta-Analysis Global Group in Chronic Heart Failure) heart failure risk score and the effect of adding natriuretic peptide for predicting mortality after discharge in hospitalized patients with heart failure. *PLoS ONE.* 2018; 13(11): e0206380. Disponible en: <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0206380>

58. Kinugasa Y, Kato M, Sugihara S, Hirai M, Kotani K, Ishida K, et al. A Simple Risk Score to Predict In-Hospital Death of Elderly Patients With Acute Decompensated Heart Failure Hypoalbuminemia as an Additional Prognostic Factor. *Circ J* 2009; 73: 2276 – 2281.
59. Lee DS, Stitt A, Austin PC, Stukel TA, Schull MJ, Chong A, et al. Prediction of Heart Failure Mortality in Emergent Care: A Cohort Study. *Ann Intern Med*. 2012;156:767-775.
60. Leong KT, Yin Wong L, Chaw Yu Aung K, Macdonald M, Cao Y, Lee S. Risk Stratification Model for 30-Day Heart Failure Readmission in a Multiethnic South East Asian Community. *Am J Cardiol*. 2017;119(9):1428-1432. Disponible: [https://www.ajconline.org/article/S0002-9149\(17\)30125-X/fulltext](https://www.ajconline.org/article/S0002-9149(17)30125-X/fulltext).
61. Lourenço P, Ribeiro A, Pintalhão M, Silva S, and Bettencourt P. Predictors of Six-Month Mortality in BNP-Matched Acute Heart Failure Patients. *Am J Cardiol*. 2015;116:744-748. Disponible: [https://www.ajconline.org/article/S0002-9149\(15\)01430-7/fulltext](https://www.ajconline.org/article/S0002-9149(15)01430-7/fulltext)
62. Martín-Sánchez FJ, Gil V, Llorens P, Herrero P, Jacob J, Fernández C, et al. Barthel Index—Enhanced Feedback for Effective Cardiac Treatment (BI—EFFECT) Study: Contribution of the Barthel Index to the Heart Failure Risk Scoring System Model in Elderly Adults with Acute Heart Failure in the Emergency Department. *JAGS*. 2012;60:493–498.
63. Miro O, Rossello X, Gil V, Martin-Sanchez FJ, Llorens P, Herrero- Puente P, et al .Predicting 30-Day Mortality for Patients With Acute Heart Failure in the Emergency Department A Cohort Study. *Ann Intern Med*. 2017;167(10):698-705.
64. Nakada Y, Kawakami R, Matsushima S, Ide T, Kanaoka K, Ueda T, et al. Simple Risk Score to Predict Survival in Acute Decompensated Heart Failure — A2B Score —. *Circ J* 2019; 83:1019–1024.
65. Okazaki H, Shirakabe A, Hata N, Yamamoto M, Kobayashi N, Shinada T, et al. New scoring system(APACHE-HF) for predicting adverse outcomes in patients with acute heart failure :Evaluation of the APACHE II and Modified APACHE II scoring systems. *Journal of Cardiology*. 2014;64:441–449.
66. Parenica J, Kala P, Goldbergova Pavkova M, Tomandl J, Spinar J, Littnerova S, et al. Natriuretic peptides, nitrite/nitrate and superoxide dismutase have additional value on top of the GRACE score in prediction of one-year mortality and rehospitalisation for heart failure in STEMI patients — Multiple biomarkers prospective cohort study. *International Journal of Cardiology*. 2016; 211:96–104.

67. Sasaki N, Lee J, Park S, Umegaki T, Kunisawa S, Otsubo T, et al. Development and Validation of an Acute Heart Failure-Specific Mortality Predictive Model Based on Administrative Data. *Canadian Journal of Cardiology* 29. 2013;29:1055-1061.
68. Scrutinio D, Ammirati E, Guida P, Passantino A, Raimondo R, Guida V, et al. Clinical utility of N-terminal pro-B-type natriuretic peptide for risk stratification of patients with acute decompensated heart failure. Derivation and validation of the ADHF/NT-proBNP risk score. *International Journal of Cardiology*. 2013;168:2120–2126.
69. Scrutinio D, Ammirati E, Guida P, Passantino A, Raimondo R, Guida V, et al. The ADHF/NT-pro-BNP risk score to predict 1-year mortality in hospitalized patients with advanced decompensated heart failure. *J HeartLungTransplant*. 2014;33:404–411.
70. Shiraishi Y, Kohsaka S, Abe T, Mizuno A, Goda A, Izumi Y, et al. Validation of the Get With The Guideline–Heart Failure risk score in Japanese patients and the potential improvement of its discrimination ability by the inclusion of B-type natriuretic peptide level. *Am Heart J*. 2016;171:33-9.
71. Shiraishi Y, Kohsaka S, Nagai T, Goda A, Mizuno A, Nagatomo Y, et al. Validation and Recalibration of Seattle Heart Failure Model in Japanese Acute Heart Failure Patients *Journal of Cardiac Failure*. 2019;25(7):561 – 567.
72. Tomcikova D, Felsoci M, Spinar J, Miklik R, Mikusova T, Jiri Vitovec, et al. Risk of in-hospital mortality identified according to the typology of patients with acute heart failure: Classification tree analysis on data from the Acute Heart Failure Database–Main registry. *Journal of Critical Care*. 2013;28:250–258.
73. Zaninović Jurjević T, Dvornik Š, Kovačić S, Matana Kaštelan Z, Brumini G, Matana A, et al. A simple prognostic model for assessing in-hospital mortality risk in patients with acutely decompensated heart failure. *Acta Clínica Bélgica*. 2018; 74: 1784-3286. Disponible en: <https://www.tandfonline.com/doi/full/10.1080/17843286.2018.1483562>

Figura 1: Flujograma de artículos seleccionados.

Tabla 1: Características poblacionales de los modelos predictivos incluidos en la revisión sistemática

Autor/año	Región	Descripción de Población	Población	Seguimiento (en meses)	Edad promedio	% Mujeres	% Diabetes	% Hipertensión arterial	% Fumador o Enfermedad pulmonar	% Enfermedad coronaria
Arenja N, et al 2011⁴⁶	Suiza	Pacientes con ICA atendidos en la emergencia del Hospital Universitario de Basilea (Suiza), durante dos períodos: mayo de 2001 a abril de 2002 y abril de 2006 a marzo de 2007.	610	12	82	46	30	71	25	-
Behnes M, et al 2016⁴⁷	Alemania	Pacientes inscritos en el estudio Mannheim NT-proBNP (MANPRO), que se realizó como estudio prospectivo controlado centrado en la Universidad Centro Médico Mannheim (UMM), Alemania.	401	8	67	49	30	67	51	32
Boralkar KA, et al 2019⁴⁸	Estados Unidos	Se usó la base de datos de la Investigación traslacional de Stanford integrada, identificando 3847 pacientes con el diagnóstico de ICA entre	580	24	76	57.6	35.5	96.7	-	59.5

		enero de 1998 y diciembre de 2016.								
Borovac JA, et al 2019³³	Croacia	Los datos se recuperaron retrospectivamente de un registro de ICA en un centro hospitalario universitario de atención terciaria durante febrero de 2010 a enero de 2013.	300	12	75.2	54.3	26.7	57.3	-	-
Cameli M, et al 2019⁴⁹	Italia	Estudio retrospectivo en una cohorte de 771 pacientes hospitalizados por signos y síntomas de ICA en la unidad de atención coronaria del Hospital Santa Maria delle Scotte en Siena, entre 2011 y 2013.	771	1, 6y 60	72.3	35	-	-	-	-
Chen YJ, et al 2017³⁸	Asia	La población se extrajo del registro de ICA del Hospital General de Veteranos de Taipei (registro HARVEST) que incluyó pacientes hospitalizados por ICA, definidos como síntomas y signos de IC de inicio reciente o que empeoran gradual o rápidamente que requieren terapia urgente.	2143	23	77	32	38.1	60.1	-	30.1

		Se incluyeron pacientes consecutivos con ICA con síntomas funcionales de clase III o IV de la New York Heart Association, presentaciones compatibles de radiografía de tórax y respuestas a diuréticos.								
Collins SP, et al 2015¹	Estados Unidos	Se reclutó un total de 2,074 sujetos con signos y síntomas de ICA del 20 de julio de 2007 al 4 de febrero de 2011, en 2 centros de atención terciaria afiliados a la universidad de Stanford y otros 2 de la comunidad.	1033	1	64	43	33	82	-	35
Delgado A, et al 2016⁵⁰	Portugal	La población total consistió en 600 pacientes hospitalizados por ICA en un servicio de cardiología del hospital Tondela Viseu de 2009 a 2011.	337	24	75.2 (80.1)	49.2 (52.1)	38.5 (32.4)	72.8 (57)	-	-
DeVore A, et al 2016²⁸	Estados Unidos	Para este análisis se utilizó la base de datos ADHERE y archivos analíticos estándar de Medicare. Es un registro grande y multicéntrico de pacientes hospitalizados con ICA en	15640	-	80.7	57.9	38.8	-	6.9	58.9

		más de 300 centros médicos de Estados Unidos. Se recopilaron datos de >185,000 pacientes entre enero de 2001 y marzo de 2006								
Diercks D, et al 2007³⁵	Estados Unidos	Se utilizó la base de datos ADHERE-EM, creada como parte del programa de registro ADHERE, y los pacientes se inscribieron entre enero de 2004 y septiembre de 2005. Se incluyeron los registros de 10 984 pacientes.	10984	-	73.1	52.2	45	79	-	-
Eulings L, et al 2014⁴³	-	Pacientes incluidos en el estudio PRIMA, un estudio prospectivo aleatorizado multicéntrico que evaluó el efecto del tratamiento de la ICC guiado por objetivos individuales NT proBNP. Se incluyeron pacientes con ICA con una concentración de NT proBNP al ingreso mínima de $\geq 1,700$ pg / mL, y los pacientes incluidos también debían demostrar una disminución en la	309	24	72	42.7	27.8	49.5	-	58.9

		concentración de NT proBNP de al menos 10% con un mínimo de 850 pg / mL durante la hospitalización.								
Formiga F, et al 2017⁵¹	España	Se utilizaron datos administrativos sobre todas las admisiones al hospital de Llobregat en un período de 18 meses (julio de 2013 diciembre de 2014) con criterios clínicos para ICA, y que fuera su primer episodio.	719	1	78.1	55.5	38.1	-	-	24.2
Garcia-Gutierrez S, et al 2016⁵²	España	Este estudio de cohorte prospectivo reclutó pacientes con síntomas de ICA que fueron atendidos en el servicio de urgencias de cualquiera de los tres hospitales del Servicio Vasco Nacional de Salud entre abril de 2011 y abril de 2013. Los pacientes eran elegibles si eran diagnosticados con IC.	1824	24	79.46	48.36	33.66	-	-	25.18
Gil V, et al 2018⁵³	España	Registro EAHFE está formado por pacientes con ICA diagnosticados en 29 centros de emergencia	1553	7 días	80	55.6	41.9	83.6	-	8.4

		españoles, en hospitales universitarios y comunitarios entre el 1 de noviembre y el 31 de diciembre de 2011								
Hsieh M, et al 2008⁵⁴	Estados Unidos	Se utilizó una muestra aleatoria generada por computadora de 5,000 pacientes dados de alta de todos los hospitales de cuidados agudos generales de Pensilvania entre los años 2003 y 2004 con diagnóstico de insuficiencia cardíaca	8384	1	75.4	56.2	38.3	-	-	-
Jacob J, et al 2015⁵⁵	España	Se utilizó el registro EAHFE, multipropósito y multicéntrico, con seguimiento prospectivo que incluye 6.597 pacientes con ICCA atendidos en 34 servicios de urgencias españoles entre 2007 y 2014.	3640	3 días	-	56.8	42.2	84.1	-	29.5
Jia Q, et al 2017⁵⁶	China	Se incluyeron pacientes con el diagnóstico de ICA en las unidades de cuidados intensivos del Departamento de Cardiología en el Hospital	2431	-	75.35 (72.27)	47.78 (49.57)	34.44 (34.62)	52.22 (56.64)	-	51.1 (51.19)

		General de China. El diagnóstico se realizó de acuerdo con los criterios de inclusión de la Sociedad Europea de Cardiología de acuerdo con los síntomas o signos, electrocardiografía, radiografía de tórax y ecocardiografía.								
Khanam SS, et al 2018⁵⁷	Corea	Se utilizó el registro de ICA coreana (KorAHF). Los pacientes que tenían signos o síntomas de IC y cumplían uno de los siguientes criterios fueron incluidos: (1) congestión pulmonar o (2) hallazgos objetivos del ventrículo izquierdo disfunción sistólica o enfermedad cardíaca estructural. Los pacientes hospitalizados por insuficiencia cardíaca aguda de 10 hospitales universitarios terciarios en todo el país se inscribieron consecutivamente desde marzo de 2011 hasta febrero de 2014.	4572	12	68.4	47	34.95	-	17.7	-

Kinugasa Y, et al 2009⁵⁸	Japón	El presente estudio incluyó a 349 pacientes mayores de 65 años que fueron hospitalizados con ICA de enero de 2004 a octubre de 2008.	349	-	79	43.9	35.5	46.9	-	-
Lee DS, et al 2012⁵⁹	Canadá	Se estudiaron datos de 12 500 pacientes que visitaron el servicio de urgencias por IC y fueron dados de alta u hospitalizados del 1 de abril de 2004 al 31 de marzo de 2007, en Ontario, Canadá.	12591	-	75.4 (75.7)	48.5 (48.4)	37.6 (38.9)	63.4 (64.2)	-	53 (50.6)
Leong K, et al 2015⁶⁰	China	Fueron seleccionados un total de 1,475 ingresos que fueron atendidos en el Hospital General de Changi entre los años 2010 y 2012	1475	1	70.6 (69.5)	41.8 (39.3)	62.3 (50.6)	74.7 (74.3)	-	65.1 (67.4)
Lim N-K, et al 2018³⁴	Corea	Se utilizó el registro de ICA coreana (KorAHF). De 2011 a 2014, 5625 pacientes consecutivos con ICA se inscribieron al ingreso. Todos los pacientes fueron evaluados para determinar si cumplían con los criterios de elegibilidad basados en	4566	12	70.3 (72.7)	48 (51.8)	40.7 (44.2)	63.8 (67.7)	17 (13.5)	29 (31.6)

		los mencionados por la Sociedad Europea de Cardiología de 2005.								
Lourenco P, et al 2015⁶¹	Portugal	Durante un período de 2 años (enero de 2009 a diciembre de 2010), se realizó un registro de ICA en el departamento de medicina interna del Hospital São João, Oporto, Portugal. Todos los pacientes ingresados con el diagnóstico primario de ICA fueron elegibles para su inclusión en el registro. Se utilizaron las pautas de la Sociedad Europea de Cardiología de 2008 para el diagnóstico de IC.	224	6	80	57	46	74	-	38
Martin-Sanchez FJ, et al 2012⁶²	España	De los 2.431 pacientes incluidos en la base de datos EAHFE, 772 no fueron seguidos porque el hospital no participaba en el estudio de resultados a corto plazo en el momento de la inclusión del paciente (n = 690) o el paciente se negó a autorizar el seguimiento (n = 82) , 139	1068	1	80	53	44	83	25	31

		eran menores de 65 años, el índice de Barthel no estaba disponible en 47, y faltaban datos esenciales para calcular la puntuación de EFECT en 407, dejando a 1.068 pacientes (44% del registro EAHFE) para su inclusión.								
Miro O, et al 2017⁶³	España	La cohorte de derivación del estudio comprende 4897 pacientes ingresados en un centro de emergencia con el diagnóstico ICA durante mayo de 2009 y noviembre-diciembre de 2011. Treinta pacientes fueron excluidos del análisis debido a la falta de seguimiento, mientras que aquellos con datos censurados fueron incluidos.	4867	1	79.9	57.1	42.2	83.4	-	29.9
Nakada Y, et al 2018⁶⁴	Japón	El presente estudio utilizó 2 cohortes de IC, el registro de IC Nara 3 (NARA-HF) 3 y el Registro japonés de IC en cardiología (JCARE-CARD). El estudio NARA-HF 3 reclutó a 1.074	739	24	73.5 (69.3)	42.4 (41.1)	-	-	-	-

		pacientes después del ingreso de emergencia por ICA entre enero de 2007 y diciembre de 2016. De la cohorte JCARE-CARD se analizaron 542 pacientes que tuvieron datos de alta y datos de seguimiento de 2 años.								
O'Connor CM, et al 2012¹²	-	El ensayo internacional PROTECT incluyó a 2033 pacientes hospitalizados con ICA y deterioro leve o moderado de la función renal. El estudio asignó al azar a pacientes con antecedentes de IC con administración de antagonista del receptor A1 rolofilina o placebo dentro de las 24 h. Los criterios de inclusión fueron: Antecedentes de ICA tratada durante al menos 14 días con terapia diurética; hospitalización por ICA que requiere terapia diurética; aclaramiento de creatinina de admisión 20–	2015	7 días	72	33	-	-	-	-

		80 ml / min; y PAS ≥ 95 mmHg.								
Okazaki H, et al 2014⁶⁵	Japón	Se recogieron datos clínicos de 824 pacientes con ICA que ingresaron en la unidad de cuidados intensivos del Hospital Chiba Hokusoh entre enero de 2000 y julio de 2012. ICA se definió como IC de nueva aparición o descompensación de la ICC con síntomas suficientes para justificar la hospitalización.	824	-	74	38	41.4	74.3	-	-
Parenica J, et al 2016⁶⁶	República Checa	Desde noviembre de 2005 hasta octubre de 2008, 913 pacientes con STEMI fueron remitidos para ICP primaria. Se incluyeron 593 pacientes con STEMI. Todos los pacientes eran caucásicos. La terapia médica estándar recomendada incluyó terapia antiplaquetaria doble, IECAs, BB y estatinas.	593	36	62 (68)	23.4 (41.8)	24.5 (50.9)	54.6 (74.5)	41.4 (18.2)	-
Sasaki N, et al 2013⁶⁷	Japón	Se recogieron datos de 19,792 pacientes en 139	8620	12	78	50.1	-	57	-	34

		hospitales con un diagnóstico primario. Los pacientes fueron incluidos en el estudio si habían sido dados de alta entre el 1 de abril de 2010 y el 31 de marzo de 2011; y habían ingresado en hospitales que habían proporcionado datos continuamente durante el período de estudio de 12 meses.									
Scrutinio D, et al 2013⁶⁸	Italia	Se analizaron a 541 pacientes ingresados por descompensación de ICC establecida con síntomas suficientes para justificar la hospitalización a la División de Cardiología de la Fundación S. Maugeri (Bari, Italia) desde abril de 2005 hasta febrero de 2011.	454	12	68.3	34	33.8	57	-	49.2	
Scrutinio D, et al 2014⁶⁹	Italia	La población del estudio comprendió 454 pacientes que cumplían los siguientes criterios de selección: hospitalización actual por empeoramiento de la ICC establecida, antecedentes de IC durante al menos 1	445	12	62	15.4	35.1	41.8	-	54.8	

	año, que reciben tratamiento crónico con terapias estándar, síntomas y evidencia funcional de clase III / IV de la Asociación Cardíaca de Nueva York de disfunción sistólica ventricular izquierda severa (fracción de eyección ventricular izquierda [FEVI] ≤ 0.30 medida por ecocardiografía bidimensional) al ingreso, y necesidad de tratamiento diurético y / o inotrópico intravenoso.								
Shiraishi Y, et al 2016⁷⁰	Japón El registro West Tokyo-Heart Failure (WET-HF) es un registro de cohortes multicéntrico amplio, continuo, prospectivo, diseñado para recopilar datos sobre antecedentes clínicos y resultados de pacientes hospitalizados por ICA. Desde enero de 2006 hasta agosto de 2014, 1.876 pacientes consecutivos de ICA de cuatro hospitales docentes	1876	-	73.3	39.4	37.1	68.1	-	-

		del área metropolitana de Tokio se inscribieron en este registro.								
Shiraiishi Y, et al 2018⁷¹	Japón	SHFM se aplicó a 2470 pacientes con ICA registrados en la base de datos de IC de West Tokyo (WET-HF) y en las bases de datos Nacional Cerebral y del centro cardiovascular de ICA (NaDEF, por sus siglas en inglés) entre los años 2006 y 2016	2470	24	74	38	38	71	-	-
Spinar J, et al 2016³⁷	República Checa	Los criterios de inclusión para la base de datos estaban en conformidad con las pautas europeas para la ICA establecidas en 2005 y fueron los siguientes: signos y síntomas de la ICA; disfunción ventricular izquierda confirmada (sistólica o diastólica); y / o respuesta positiva a la terapia.	5846	33	74	43.3	44	75	-	24.3
Thorvaldson T, et al 2017³⁶	Estados Unidos	Se utilizó el registro ARIC, el cual contiene información de las hospitalizaciones de 21	1852	-	77	65	-	-	-	-

		hospitales en 4 estados de Estados Unidos. La muestra de derivación comprendió 1852 hospitalizaciones de 2005 a 2011 y la muestra de validación comprendió 821 hospitalizaciones de 2012 a 2013.								
Tomcikova D, et al 2013⁷²	República Checa	El análisis actual se basa en los datos de 4153 registros de pacientes entre septiembre de 2006 y octubre de 2009. El análisis involucra solo la primera hospitalización por ICA para cada paciente durante el período de estudio. La cohorte analizada también incluyó a pacientes hospitalizados por ICA en un estudio piloto del mayor centro del Hospital Universitario de Brno, que comenzó la inscripción antes, en diciembre de 2004.	4153	-	73.9 (73.5)	42.6 (42.2)	41.7 (40.5)	70.7 (71.6)	15.9 (16.4)	-
Yagyu T, et al 2017⁴⁴	Japón	Entre enero de 2009 y diciembre de 2013, se inscribieron un total de	1035	-	77.5	47.6	30.3	64.6	-	29.8

		1.062	pacientes consecutivos con ICA ingresados en el hospital de Tonoyaka. Se realizó un diagnóstico clínico de ICA basado en la presentación clínica, la evaluación de la ecocardiografía y las pruebas de laboratorio de péptidos natriuréticos, y se confirmó adicionalmente mediante registros clínicos basados en los criterios de Framingham para el diagnóstico de IC.								
Zaninovic – Jurjevic T, et al 2019⁷³	Croacia	El grupo de derivación se estableció retrospectivamente de los pacientes tratados por ICA en el departamento de cardiología del hospital Rijeka durante el período de tres años. En el grupo de validación hubo un total de 591 pacientes con ICA hospitalizados consecutivamente durante un año de seguimiento. Los criterios de inclusión para el grupo de derivación	654	12	74.2 (76.4)	58 (62.5)	38 (45.5)	28 (13)	-		59.5 (64.5)

fueron: diagnóstico primario de admisión de ICA (ya sea con fracción de eyección conservada, de rango medio o reducida), pacientes mayores de 18 años y la existencia de hallazgos bioquímicos comunes básicos dentro de las 24 horas posteriores al ingreso.

BB: Beta bloqueadores IC: Insuficiencia Cardíaca ICA: Insuficiencia Cardíaca Aguda ICC: Insuficiencia Cardíaca Crónica ICCA: Insuficiencia Cardíaca Crónica Agudizada IECA: Inhibidor de la enzima convertidora de angiotensina PAS: Presión Arterial Sistólica STEMI: Infarto de Miocardio con segmento ST eleva

Tabla 2: Características de diseo de los modelos predictivos incluidos en la revisión sistemática

Autor/año	Diseño de artículo	de Variables Exposición	Estrategia de selección	Método de Modelo	de Tipo de Modelo	de Validación
Arenja N, et al 2011⁴⁶	Cohorte Prospectivo	NUS, PAS, Pcr	A través del árbol de clasificación y regresión	Análisis tipo árbol	Desarrollo	Interna
Behnes M, et al 2016⁴⁷	análisis post hoc de un estudio	Edad, Sexo, Clasificación NYHA, FEVI, Na,	Basada en variables con	Regresión de cox	Desarrollo	Interna

	prospectivo controlado	Hb, IECA/ARA, BB, NT-proBNP, Gremlina, Osteopontina, Galectina	asociación en artículos previos				
Boralkar KA, et al 2019⁴⁸	Cohorte Retrospectivo	PAS, NUS, Na, Edad, FC, Raza negra, EPOC	Basada en asociación con las variables de admisión	Análisis de riesgos proporcionales de Cox	Desarrollo	Interna	
Borovac JA, et al 2019³³	Cohorte Retrospectivo	PAS, Na, CP, Pcr, FEVI, HP, AcU, TFGE	Variables con rol pronóstico según la literatura	Regresión de cox	Desarrollo	Interna	
Cameli M, et al 2019⁴⁹	Cohorte Retrospectivo	Edad, Pcr, VNI, ACV o ITA, FEVI, HP, DVM	Selección de variables según asociación con la mortalidad	Modelo lineal Generalizado con respuesta de Poisson	Desarrollo	Interna	
Chen YJ, et al 2017³⁸	Cohorte Prospectivo	FA, Hb, Edad avanzada, Alteración de parámetros renales, DM (1) + AcU (2)	Variables del estudio original de la escala AHEAD y ácido úrico	Regresión de cox	Desarrollo (1) Validación (2)	Interna (1) Externa (2)	
Collins SP, et al 2015¹	Cohorte Prospectivo Observacional	Edad, IMC, PNC, PAD, NUS, Na, FR, SatO2, Troponina I, diálisis, IECA, Duración de QRS	Factores de riesgo en estudios previos	Regresión spline	Desarrollo	Interna	
Delgado A, et al 2016⁵⁰	Cohorte Retrospectivo	Edad, Proporción E/e', PNC, U, Na, Sin indicación de IECA/ARA al alta	Según su asociación con la mortalidad de la	Regresión logística y de cox	Desarrollo	Interna	

			población del estudio				
DeVore A, et al 2016²⁸	Cohorte Prospectivo	Edad, Pulso, PAS, FEVI, PNC, Troponina, Na, NUS, Pcr	Variables seleccionadas en estudios previos (ADHERE)	Regresión logística	Desarrollo y validación	Interna y externa	
Diercks D, et al 2007³⁵	Revisión retrospectiva	NUS, PAS, Pcr	Variables seleccionadas en estudios previos (ADHERE)	Regresión logística	Validación	Externa	
Eulings L, et al 2014⁴³	Estudio multicéntrico prospectivo aleatorizado	NT-proBNP	Selección de variable según literatura	Regresión multivariable de cox	Validación	Externa	
Formiga F, et al 2017⁵¹	Cohorte Retrospectivo	Edad, género, IC previa, EAor, ACV, EPOC, Demencia, PAS, FC, FR, Na, NUS, Hct, Pcr, Glucosa, FEVI.	NR	Regresión logística	Validación	Externa	
Garcia-Gutierrez S, et al 2016⁵²	Cohorte Prospectivo	EP, HP-EMG en los últimos 2 años, Glucosa, NUS	Variables asociadas a la mortalidad en el artículo	Regresión Logística multivariada	Desarrollo y validación	Interna y externa	
Gil V, et al 2018⁵³	Cohorte Retrospectivo	Edad, EMG, Cáncer activo, Metolazona, PAS, FC, satO2, Pcr, K, Troponina	Variables de Estudios anteriores (EHMGR)	NR	Validación	Externa	
Hsieh M, et al 2008⁵⁴	Cohorte Retrospectivo	Género, IMA, Historia de angina,	NR	NR	Desarrollo y validación	Interna y Externa	

		Historia previa de angioplastia, DM, Enfermedad pulmonar, Pulso, FR, PAS, Pcr, Temperatura, Na, NUS, GB, pH arterial, EKG de: IMA, Isquémica cardiaca, Radiografía: Congestión pulmonar y efusión pleural					
Jacob J, et al 2015⁵⁵	Cohorte prospectivo multicéntrico	Edad, Clasificación NYHA, PAS, satO2, Na, Uso de Inotrópicos o vasopresores, VNI	NR	Regresión logística	Desarrollo y validación	y	Interna y Externa
Jia Q, et al 2017⁵⁶	Cohorte Retrospectivo	TFGE, FC, FEVI, NT-proBNP, pH arterial	Variables con asociación según estudios previos	Regresión logística	Desarrollo y validación	y	Interna y Externa
Khanam SS, et al 2018⁵⁷	Cohorte Prospectivo	Edad, Género, IMC, Fumador, PAS, DM, Clasificación NYHA, FEVI, EPOC, ICA >18 meses, Pcr, BB, IECA o ARA, PNC o NT-proBNP	Variables de escala previa (MAGGIC)	Regresión logística	Desarrollo y validación	y	Interna y Externa

Kinugasa Y, et al 2009⁵⁸	Cohorte Retrospectivo	HP, Na, Albúmina, NUS, PNC	NR		Regresión logística multivariable	Desarrollo	Interna
Lee DS, et al 2012⁵⁹	Cohorte Retrospectivo	Edad, EMG, Cáncer activo, Metolazona, PAS, FC, satO2, Pcr, K, Troponina	NR		Regresión logística multivariable	Desarrollo y validación	Interna y Externa
Leong K, et al 2015⁶⁰	Cohorte Retrospectivo Observacional	Número de HP, Tiempo de estadia Hospitalaria, Pcr, NT-proBNP, Duración de QRS, Número de referencias médicas, indicación de BB al alta	Variables que mostraron asociación según la regresión		Regresión Logística	Desarrollo	Interna
Lim N-K, et al 2018³⁴	Cohorte Prospectivo	Edad, clasificación NYHA, HTA, EPOC, etiología (cardiopatías previas), PAS al alta, FEVI, BB al alta, IECA, Na, PNC	Variables que mostraron asociación según la regresión		Regresión Logística Múltiple	Desarrollo	Interna
Lourenco P, et al 2015⁶¹	Estudio de casos y control	1er modelo: FA, HTA, FC de ingreso, PAS de ingreso, Hb, U, Albuminas, BB o IECA/ARA, PNC, 2do modelo: + FEVI, NYHA IV al ingreso, ICA	Variables asociadas con el análisis univariado y clásicamente relaciones a la ICA		Regresión de cox en ambos modelos	Desarrollo	Interna

		Antecedente de enfermedad isquémica					
Martin-Sanchez FJ, et al 2012⁶²	Cohorte prospectivo Multicéntrico	NR	Variables incluidas en modelos previos (EFFECT y BI)	Análisis tipo árbol	Desarrollo	Interna	
Miro O, et al 2017⁶³	Cohorte Prospectivo	Puntaje de Índice de Barthel al ingreso, PAS, Edad, NT-proBNP, K, Troponina, NYHA IV al ingreso, FR, satO2, SICA, Hipertrofia en EKG, Pcr	Variables que encontraron asociación a través de la regresión	Regresión logística	Desarrollo y validación	Interna y externa	
Nakada Y, et al 2018⁶⁴	Cohorte Prospectivo	Edad, Hb, PNC	Variables que encontraron asociación en estudios previos	Regresión multivariable de cox	Desarrollo y validación	Interna y externa	
O'Connor CM, et al 2012¹²	Estudio aleatorizado controlado con placebo	NUS, PAS, FR, HP	Variables seleccionadas en al menos 20 bases de datos previos.	Regresión multivariable de cox	Desarrollo y validación	Interna y externa	
Okazaki H, et al 2014⁶⁵	Cohorte Retrospectivo	Edad, PAM, Pulso, Na, K, Pcr, Htco, EG	Variables utilizadas en la escala APACHE II	Regresión multivariable de cox	Desarrollo	Interna	

Parenica J, et al 2016 ⁶⁶	Cohorte Prospectivo	Edad, FC, PAS, Pcr al ingreso, Clasificación Killip, Paro cardíaco al ingreso, elevación de segmento ST, Elevación de marcadores cardíacos	Variables asociadas a la mortalidad en el artículo	Regresión logística multivariable	Validación	Externa
Sasaki N, et al 2013 ⁶⁷	Cohorte Retrospectivo	Modelo 1: Género, edad, Vía de ingreso, Falla respiratoria severa por ICA, Enfermedad isquémica cardíaca, HTA, FA, Arritmia grave, ERC, Shock Modelo 2: + clasificación de NYHA	Variables seleccionadas en literatura previa	Regresión logística multivariable	Desarrollo y validación	Interna y externa
Scrutinio D, et al 2013 ⁶⁸	Cohorte Retrospectivo	EPOC, PAS, TFGE, Na, NT-proBNP, FEVI, Regurgitación tricúspidea severa	Variables que mostraron asociación en la regresión	Regresión logística múltiple	Desarrollo y validación	Interna y externa
Scrutinio D, et al 2014 ⁶⁹	Cohorte Retrospectivo	EPOC, PA, TFGE, Na, Hb, NT-proBNP, FEVI, Regurgitación tricúspidea severa	Variables utilizadas en un modelo previo NT-proBNP score	Regresión de cox	Validación	Externa

Shiraishi Y, et al 2016⁷⁰	Cohorte Prospectivo Multicéntrico	Raza, edad, PAS, FC, TFGE, NUS, Na, EPOC	Variables utilizadas en un modelo previo GWTG-HF risk score	Regresión logística multivariable	Validación	Externa
Shiraiishi Y, et al 2018⁷¹	Cohorte Prospectivo	EPOC, PA, TFGE, Na, Hb, NT-proBNP, FEVI, Regurgitación tricúspide severa	Variables utilizadas en artículos anteriores	Regresión de cox	Validación	Externa
Spinar J, et al 2016³⁷	Cohorte Prospectivo Multicéntrico	FA, Hb, Edad avanzada, Alteración de parámetros renales, DM	A través de modelo de regresión multivariado y expertos	Regresión multivariada de cox	Desarrollo y validación	Interna y externa
Thorvaldsen T, et al 2017³⁶	Cohorte Retrospectivo	Edad, PAS, NUS, Na, ACV, EPOC, Hb	Variables utilizadas en el modelo EFFECT y a través de regresión logística	Regresión logística	Desarrollo y validación	Interna y externa
Tomcikova D, et al 2013⁷²	Cohorte Prospectivo	Shock Cardiogénico, Pcr, PAS	A través del árbol de clasificación y regresión	Análisis tipo árbol	Desarrollo y validación	Interna y externa
Yagyu T, et al 2017⁴⁴	Cohorte Retrospectivo	NUS, Administración de catecolaminas, Plaquetas, Bilirrubina, PCR	Variables utilizadas en un modelo previo (escala GWTG-HF)	Regresión logística multivariada	Desarrollo y validación	Interna y externa

Zaninovic – Cohorte Jurjevic T, et Retrospectivo al 2019⁷³	Edad, PAS, PAD, Variables FC, Na, Pcr, U, AcU, seleccionadas en HTA, ACV o ITA, modelos previos Bloqueo de haz izquierdo, Anemia severa, NT-proBNP	Curvas ROC	Desarrollo y Interna y externa validación
--	---	------------	--

AcU: Ácido Úrico ACV: Accidente Cerebro Vascular ARA: Antagonista de los receptores de angiotensina BB: Beta – Bloqueadores
DVM: Disfunción de válvula mitral DM: Diabetes Mellitus DVM: Disfunción de válvula mitral EAor: Estenosis Aórtica EG: Escala
de Glasgow EMG: Ingreso por emergencia por insuficiencia cardíaca aguda EPOC: Enfermedad Pulmonar Obstructiva Crónica ERC:
Enfermedad Renal Crónica FA: Fibrilación auricular FC: Frecuencia Cardíaca FE VI: Fracción de Eyección Ventricular Izquierda
FR: Frecuencia respiratoria Hb: Hemoglobina Htco: Hematocrito HP: Hospitalización previa por insuficiencia cardíaca aguda HTA:
Hipertensión arterial ICA: Insuficiencia Cardíaca Aguda IECA: Inhibidor de la enzima convertidora de angiotensina IMA: Infarto de
miocardio Agudo IMC: Índice de Masa Corporal ITA: Isquemia Transitoria Aguda K: Potasio Na: Sodio sérico NT-proBNP: N-
terminal del pro-péptido natriurético tipo B NUS: Nitrógeno ureico en sangre NYHA: New York Heart Association Pcr: Creatinina
Sérica Proporción E/e': Medida ecográfica que mide la relación entre la velocidad de entrada mitral temprana y la velocidad anular temprana
diastólica mitral PA: Presión Arterial PAD: Presión Arterial Diastólica PAM: Presión Arterial Media PAS: Presión Arterial Sistémica
CP: Creatinphosfokinasa PCR: Proteína C reactiva Sato2: Saturación de oxígeno SICA: Síndrome coronario Agudo TFGE: Tasa
de Filtración Glomerular Efectiva U: Urea VNI: Ventilador no invasivo

Tabla 3: Estadísticas de desempeño de los modelos predictivos incluidos en la revisión sistemática.

Autor	ROC (CI 95%)	¿Cali braci ón repor tada?	Hosmer- Lemeshow p value	O:E statistic and 95%CI	¿Discrimin ación reportada?	¿Benefici o Neto reportado ?	¿Reclas ificació n Reporta do?	IDI 95%CI	and	NRI 95%CI	and	Desenlace
--------------	---------------------	-----------------------------------	---------------------------------	--------------------------------	------------------------------------	-------------------------------------	---------------------------------------	------------------	------------	------------------	------------	------------------

Arenja N, et al 2011⁴⁶	.	No	No	No	No	No	No	No	No	Mortalidad en 1 año
Boralkar KA, et al 2019⁴⁸	0.72	No	No	No	Yes	No	Yes	GWTG-HF +IMRS: 2.14% (0.99 - 3.30) Subgroup with NT-proBNP: GWTG-HF +IMRS: 1.64%(0.49 - 2.78) +NT-proBNP: 3.15% (1.52- 4.78) +IMRS and NT-proBNP: 4.03%(2.13 - 5.93) IMRS+NT-proBNP: 3.26% (1.62 - 4.90)	GWTG-HF +IMRS: 36.2% (17.5 - 54.9) Subgroup with NT-proBNP: GWTG-HF +IMRS: 25.8% (3.63 - 49.9) +NT-proBNP: 36.3% (14.4 - 58.3) +IMRS and NT-proBNP: 39.9% (18.0 - 61.8) IMRS +NT-proBNP: 46.2% (24.4 - 68.0)	Mortalidad en 3 años
Borovac JA, et al 2019³³	0.907(0.867-0.939)	Yes	0.14	No	Yes	No	No	No	No	Mortalidad en 1 año
Cameli M, et al 2019⁴⁹	30 días: 0.78	No	No	No	Yes	No	No	No	No	Mortalidad en 30 días, 6 meses y 5 años

	6 meses: 0.79, 5 años: 0.76.										
Chen YJ, et al 2017³⁸	.	No	No	No	No	No	No	No	No	vs AHEAD-U: <i>TCM</i> - <i>SC</i> ⁺ : 19,66% <i>TCM</i> - <i>CC</i> ^o : 10,22% <i>MC</i> - <i>SC</i> ^o : 20,08% <i>MC</i> - <i>CC</i> ⁻ : 10,25%,	Mortalidad en 3 años
Diercks D, et al 2007³⁵	.	No	No	No	No	No	No	No	No	No	Mortalidad en 2 años
Eulings L, et al 2014⁴³	0.85 (0.81- 0.90)	Yes	0.77	No	Yes	No	No	+ NT-proBNP: 6% - 13%	No	No	Mortalidad en 90, 180 y 365 días
Gil V, et al 2018⁵³	0.741(0.688- 0.793)	No	No	No	Yes	No	No	No	No	No	Mortalidad en 7 días
Jacob J, et al 2015⁵⁵	Derivación: 0.8(0.76- 0.84) Validación: 0.76 (0.7- 0.82)	No	No	No	Yes	No	No	No	No	No	Mortalidad en 3 días

Khanam SS, et al 2018⁵⁷	0.86	No	No	No	Yes	No	No	No	No	Mortalidad en 1 año
Kinugasa Y, et al 2009⁵⁸	Derivación: 0.807 Validación: 0.806	Yes	0.935	No	Yes	No	No	No	No	Mortalidad posterior al alta en 6 meses, 1 año o 5 años
Lee DS, et al 2012⁵⁹	.	No	No	No	no	No	No	No	No	Mortalidad en 7 días
Lourenco P, et al 2015⁶¹	EFFECT : 0.69 (0.62-0.76) EFFECT -BI: 0.754(0.69-0.81)	No	No	No	Yes	No	No	No	No	Mortalidad en 6 meses
Martin-Sanchez FJ, et al 2012⁶²	Derivación: 0.836(0.818-0.853) Validación: 0.828 (0.802-0.853)	Yes	0.99	No	Yes	No	No	No	No	Mortalidad en 30 días
Miro O, et al 2017⁶³	0.711(0.671-0.751)	No	No	No	Yes	No	No	Grupo 7 vs. 1 5.8% (5.2–6.9) < Grupo 7 vs. 2	Grupo 7 vs. 1 48% (40–55.9) Grupo 7 vs. 2	Mortalidad en 30 días

								6.2% (5.3–7.1)	50% (42.1–58)		
								Grupo 7 vs. 3	50.2% (42.3–58.1)		
								Grupo 7 vs. 4	36.4% (28.5–44.3)		
								Grupo 7 vs. 5	30.6% (22.7–38.5)		
								Grupo 7 vs. 6	41% (33.1–48.9)		
Nakada Y, et al 2018⁶⁴	0.67	Yes	0.67	No	Yes	No	No	No	No	No	Mortalidad en 2 años, fuera del hospital.
Scrutinio D, et al 2013⁶⁸	Derivación: 0.839(0.798–0.88) Validación: 0.768 (0.711–0.817)	Yes	Derivación: 0.542 Validación: 0.251	No	Yes	No	Yes	+ proBNP: 3.7% (1.6–5.7)	NT- + NT-proBNP: 12.9% (4.5–21.3)		Mortalidad en 1 año
Shiraiishi Y, et al 2018⁷¹	0.75	No	No	No	Yes	No	No	No	No	No	Mortalidad 1 año posterior al alta

Spinar J, et al 2016³⁷	0.72	No	No	No	Yes	No	No	No	No	Mortalidad en 1 año
Thorvaldsen T, et al 2017³⁶	Derivación: 0.76 28 días: 0.72 1 año: 0.72 Validación: 0.73 28 días: 0.73 1 año: 0.71	Yes	28 días: 0.48 1 año: 0.33	No	Yes	No	No	No	No	Mortalidad en 28 días o 1 año
Behnes M, et al 2016⁴⁷	.	Yes	Gremlin-1: 0.9951 <i>RH-TP</i> 1 año [†] Osteopontina: 0.9905 <i>TCM</i> - <i>PICA</i> 5 años** Galectin-3: 0.5775 <i>TCM-TP</i> 1 año [^] .	No	Yes	No	Yes	Gremlin-1: 8.04% (1.2 - 14.8) <i>RH-PICA</i> 1 año*, Osteopontina: 2.68% (-0.39 - +5) <i>TCM</i> - <i>PICA</i> 5 años**, Galectin-3: 5.7% (2 - 8) <i>TCM-TP</i> 5 años [†] .	Gremlin-1: 58.81% (16-100) <i>RH-PICA</i> 1 año*, Osteopontina: 33.48% (-4 - +71) <i>TCM</i> - <i>PICA</i> 5 años**, Galectin-3: 49,82% (27 - 71) <i>TCM-TP</i> 5 años [†] .	Mortalidad en 1 y 5 años. Reingreso Hospitalario en 1 y 5 años.

Delgado A, et al 2016 ⁵⁰	0.75	Yes	No	No	Yes	No	No	No	No	Mortalidad o Reingreso Hospitalario en 6, 12 y 24 meses
Lim N-K, et al 2018 ³⁴	.	No	No	No	no	No	No	No	No	Mortalidad o Reingreso hospitalario en 30 días
Okazaki H, et al 2014 ⁶⁵	0.84(0.78-0.9)	No	No	No	Yes	No	No	+ BNP : 3.4% (0.9-6) + NT-proBNP : 3.8% (0.2-7.3) + SOD : 5.4% (3.4-7.5) + NOx : -0.4% (-3.5 - 2.6)	+ BNP : 72.1% (45.5-98.7) + NT-proBNP : 106.6%(80.2-132.9) + SOD : 85.2% (58.7-100) + NOx : 53.6% (19.8-87.4)	Mortalidad o reingreso hospitalario en 90 días
Parenica J, et al 2016 ⁶⁶	0.863	No	No	No	Yes	No	No	No	No	Mortalidad o reingreso hospitalario en 1 año
Garcia-Gutierrez S, et al 2016 ⁵²	0.833(0.788-0.878)	No	No	No	Yes	No	No	No	No	Mortalidad o exacerbación de síntomas en 7 días
Hsieh M, et al 2008 ⁵⁴	.	No	No	No	no	No	No	No	No	Mortalidad o exacerbación de síntomas

												posterior al alta en 30 días
DeVore A, et al 2016 ²⁸	Derivación: 0.74 Validación: 0.72 Validación Externa: 0.63	Yes	Validación Interna: 0.14 Validación Externa: 0.001	Derivación: 0.43 Validación: 0.14	Yes	No	No	No		No		Exacerbación intrahospitalaria
Scrutinio D, et al 2014 ⁶⁹	Mortalidad acumulada: 0.738 Mortalidad post alta: 0.741	No	No	No	Yes	No	No	No		No		Mortalidad intrahospitalaria o 1 año posterior al alta
Sasaki N, et al 2013 ⁶¹	Modelo 1: 0.76 (0.74-0.78) Modelo 2: 0.80(0.78-0.82)	Yes	Modelo 1: 0.44 Modelo 2: 0.88	No	Yes	No	No	No		No		Mortalidad intrahospitalaria
Jia Q, et al 2017 ⁵⁶	Maggic: 0.734(0.	Yes	0.5559	No	Yes	No	Yes	+ proBNP: 0.2%	NT-	+ NT-proBNP: 33.40%		Mortalidad Intrahospitalaria

	72-0.749).										
	Maggic + BNP o NTproB NP: 0.736(0.721-0.750)										
Shiraishi Y, et al 2016⁷⁰	0.763 (0.7-0.826) +BNP: 0.818 (0.771-0.865)	Yes	0.577	No	Yes	No	No	No	No	No	Mortalidad intrahospitalaria
Tomcikova D, et al 2013⁷²	Derivación: 0.823 (0.792-0.854) Validación: 0.832(0.789-0.876)	No	No	No	Yes	No	No	No	No	No	Mortalidad intrahospitalaria

Yagyu T, et al 2017⁴⁴	Derivación: 0.888 (0.837-0.938) Validación: 0.902(0.858-0.945)	Yes	Derivación: 0.642 Validación: 0.648	No	Yes	No	No	No	No	Mortalidad intrahospitalaria
Zaninovic – Jurjevic T, et al 2019⁷³	Derivación: 0.741(0.701-0.776) Validación: 0.741 (0.706-0.774)	No	No	No	Yes	No	No	No	No	Mortalidad intrahospitalaria
Collins SP, et al 2015¹	0.68 (0.63-0.74)	Yes	No	No	Yes	No	No	No	No	Mortalidad, exacerbación de síntomas o Reingreso Hospitalario en 30 días
O'Connor CM, et al 2012¹²	0.779	No	No	No	Yes	No	No	No	No	Mortalidad, Exacerbación de síntomas o Reingreso hospitalario en 7 días

Formiga F, et al 2017⁵¹	30 días [§] -0.649(0.57-0.72) 90 días [§] -0.621(0.560–0.681)	No	No	No	Yes	No	No	No	No	No	Reingreso Hospitalario en 30 y 90 días
Leong K, et al 2015⁶⁰	Modelo Básico: 0.647(0.62-0.674) Modelo Clínico: 0.679(0.653-0.705) Modelo Complejo: 0.71(0.685-0.735)	Yes	0.3826	No	Yes	No	No	+ Na<135 mmol/L + BNP≥700 pg/mL or NTproBNP≥8,000 pg/mL: 2% (1-3)	+ Na<135 mmol/L + BNP≥700 pg/mL or NTproBNP≥8,000 pg/mL Con Categorización : 17% (11-24) Sin Categorización : 42% (33-51)	Reingreso hospitalario en 30 días	

‡ RH-TP 1 año: Re-Hospitalización en todos los pacientes en un plazo de un año. *RH-PICA 1 año: Re-Hospitalización en Pacientes con Insuficiencia Cardíaca Aguda en un plazo de 1 año. **TCM-PICA 5 años: Toda Causa de Mortalidad en Pacientes con Insuficiencia Cardíaca Aguda en un plazo de 5 años. †TCM-TP 5 años: Toda Causa de Mortalidad en Todos los Pacientes del estudio en un plazo de 5 años. ^TCM - TP 1 año: Toda Causa de Mortalidad en Todos los Pacientes del estudio en un plazo de 1 año. °TCM - SC: Toda Causa de Mortalidad Sin Categorización de riesgo de grupos poblacionales. °TCM – CC: Toda Causa de Mortalidad Con Categorización de riesgo de grupos poblacionales.

°MC – SC°: Mortalidad por Causa Cardiovascular Sin Categorización de riesgo de los grupos poblacionales. °MC - CC: Mortalidad por Causa Cardiovascular Con Categorización de riesgo de los grupos poblacionales. §*Riesgo de Reingreso hospitalario Grupo 1 (Edad), Grupo 2 (Hb), Grupo 3(BNP), Grupo 4 (Edad+Hb), Grupo 5 (Edad+BNP), Grupo 6 (Hb+BNP), Grupo 7 (Edad+Hb+BNP)* BNP: Péptido Natriurético tipo B ; NT-proBNP: N-terminal de la prohormona natriurética de péptido tipo B; SOD: Superóxido Dismutasa; Nox:nitrito/ nitrato. IDI : integrated discrimination improvement; NRI: net reclassification improvement; cNRI: continuous net reclassification improvement; CI: Intervalo de Confianza ; Hb :Hemoglobina. (1) AHEAD-U (2) AHEAD.

Tabla 4: Resultados de evaluación de riesgo de sesgo utilizando la herramienta PROBAST.

Estudios	Riesgo de Sesgo				Aplicabilidad			General	
	Participantes	Predictores	Outcome	Análisis	Participantes	Predictores	Outcome	RDS	Aplicabilidad
1. Behnes M, et al 2016⁴¹	+	+	+	+	+	+	+	+	+
2. Borovac JA, et al 2019²⁷	+	+	+	+	+	+	+	+	+
3. Cameli M, et al 2019⁴³	+	+	+	+	+	+	+	+	+
4. Delgado A, et al 2016⁴⁴	+	+	+	+	+	+	+	+	+
5. Diercks D, et al 2007²⁹	+/+	+/+	+/+	+/+	+/+	+/+	+/+	+	+
6. Garcia-Gutierrez S, et al 2016⁴⁶	+	+	+	+	+	+	+	+	+
7. Gil V, et al 2018⁴⁷	+/+	+/+	+/+	+/+	+/+	+/+	+	+	+
8. Kinugasa Y, et al 2009⁵²	+	+	+	+	?	+	+	+	+
9. Lee DS, et al 2012⁵³	+	+	?	-	+	+	+	+	+
10. Lourenco P, et al 2015⁵⁵	+	+	+	+	+	+	+	+	+
11. Nakada Y, et al	+/+	+/+	+/+	+/+	+/+	+/+	+/+	+	+

2018⁵⁸									
12. O'Connor CM, et al 2012⁹	+/+	+/+	+/+	+/+	+/+	+/+	+/+	+	+
13. Okazaki H, et al 2014⁵⁹	+/+	+/+	+/+	+/+	+/+	+/+	+/+	+	+
14. Parenica J, et al 2016⁶⁰	+	+	+	+	+	+	+	+	+
15. Sasaki N, et al 2013⁶¹	+	+	+	+	+	+	+	+	+
16. Scrutinio D, et al 2014⁶³	+/+	+/+	+/+	+/+	+/+	+/+	+/+	+	+
17. Shiraishi Y, et al 2018⁶⁵	-	+	+	+	-	+	+	+	+
18. Thorvaldsen T, et al 2017³⁰	+/+	+/+	+/+	+/+	+/+	+/+	+/+	+	+
19. Tomcikova D, et al 2013⁶⁶	+/+	+/+	+/+	+/+	+/+	+/+	+/+	+	+
20. Yagy T, et al 2017³⁸	+/+	+/+	+/+	+/+	+/+	+/+	+/+	+	+
21. Zaninovic – Jurjevic T, et al 2019⁶⁷	+/+	+/+	+/+	+/+	+/+	+/+	+/+	+	+
22. Boralkar KA,	+	+	+	+	+	?	+	+	-

et al 2019⁴²									
23. Scrutinio D, et al 2013⁶²	+/+	+/+	+/+	+/+	+/+	?/?	+/+	+	-
24. Arenja N, al 2011⁴⁰	+	+	?	-	+	+	+	-	+
25. Chen YJ, et al 2017³² (AHEAD)	-	+	+	-	+	+	+	-	+
26. Chen YJ, et al 2017³² (AHEAD-U)	-	+	+	-	+	+	+	-	+
27. Collins SP, et al 2015¹	-	+	+	-	+	+	+	-	+
28. Eulings L, et al 2014³⁷	+	+	+	-	+	+	+	-	+
29. Formiga F, et al 2017⁴⁵	+	+	+	-	+	+	+	-	+
30. Hsieh M, et al 2008⁴⁸	?	+	+	-	+	+	+	-	+
31. Jia Q, et al 2017⁵⁰	+/+	+/+	+/+	-/-	+/+	+/+	+/+	-	+
32. Khanam SS, et al 2018⁵¹	+/+	+/+	?/?	-/-	+/+	+/+	+/+	-	+

33. Leong K, et al 2015⁵⁴	+/+	+/+	+/+	-/-	+/+	+/+	+/+	-	+
34. Lim N-K, et al 2018²⁸	-	+	+	+	-	+	+	-	+
35. DeVore A, et al 2016²²	+	+	+	-	+	-	+	-	-
36. Martin-Sanchez FJ, et al 2012⁵⁶	+	+	+	-	+	-	+	-	-
37. Jacob J, et al 2015⁴⁹	+	+	+	?	+	+	+	?	+
38. Miro O, et al 2017⁵⁷	+	+	+	?	+	+	+	?	+
39. Shiraishi Y, et al 2016⁶⁴	+	-	+	+	+	-	+	?	+
40. Spinar J, et al 2016³¹	?	+	+	?	?	+	+	?	?

+ indica bajo RDS / baja preocupación con respecto a la aplicabilidad; - indica un alto RDS / alta preocupación con respecto a la aplicabilidad; y ? indica poro claro RDS / preocupación poco clara con respecto a la aplicabilidad.

*El artículo desarrollado por Chen YJ, et al; cuenta con 2 análisis de Probast, pues contaba con dos modelos con outcomes diferentes. El primero buscaba validar externamente el modelo AHEAD, y el segundo creaba una nueva escala llamada AHEAD-U con su respectiva validación internas

ANEXOS

Nº1:

PubMed search strategy (short form)

“acute heart failure” AND (“clinical prediction” OR “clinical model*” OR “clinical score*” OR “72ecisión rule*” OR “predictive outcome*” OR “predictive rule*” OR “predictive score*” OR “predictive value” OR “predictive risk*” OR “prediction outcome*” OR “prediction rule*” OR “prediction score*” OR “prediction value*” OR “prediction risk*” OR “risk assessment” OR “risk score*” OR “validation 72ecisión*” OR “validation rule*” OR “validation score*” OR (derivation AND validation) OR (sensitivity AND specificity))

PubMed search strategy (long form)

“acute heart failure”[All Fields] AND (“clinical prediction”[All Fields] OR “clinical model*”[All Fields] OR “clinical score*”[All Fields] OR “72ecisión rule*”[All Fields] OR “predictive outcome*”[All Fields] OR “predictive rule*”[All Fields] OR “predictive score*”[All Fields] OR “predictive value”[All Fields] OR “predictive risk*”[All Fields] OR “prediction outcome*”[All Fields] OR “prediction rule*”[All Fields] OR “prediction score*”[All Fields] OR “prediction value*”[All Fields] OR “prediction risk*”[All Fields] OR “risk assessment”[All Fields] OR “risk score*”[All Fields] OR (validation[All Fields] AND (“Decision (Wash D C)”[Journal] OR “72ecisión”[All Fields])) OR (validation[All Fields] AND (rule[All Fields] OR rule’[All Fields] OR rule’’[All Fields] OR rule’per[All Fields] OR rule’s[All Fields] OR rule1[All Fields] OR rule14[All Fields] OR rule2[All Fields] OR rule3[All Fields] OR rule4[All Fields] OR rule77[All Fields] OR rulea[All Fields] OR ruleable[All Fields] OR rulebase[All Fields] OR rulebased[All Fields] OR rulebases[All Fields] OR rulebe[All Fields] OR rulebender[All Fields] OR rulebook[All Fields] OR rulebooks[All Fields] OR rulebound[All Fields] OR rulebreaking[All Fields] OR rulec[All Fields] OR rulect[All Fields] OR ruled[All Fields] OR ruled’[All Fields] OR ruledynamics[All Fields] OR ruleed[All Fields] OR rulefit[All Fields] OR rulefit3[All Fields] OR rulefor[All Fields] OR rulefr[All Fields] OR ruleful[All Fields] OR rulefully[All Fields] OR rulego[All Fields] OR rulei[All Fields] OR rulein[All Fields] OR ruleless[All Fields] OR rulelike[All Fields] OR rulem[All Fields] OR rulemake[All Fields] OR

rulemakers[All Fields] OR rulemaking[All Fields] OR rulemaking's[All Fields] OR rulemakings[All Fields]
 OR rulemanager[All Fields] OR ruleminer[All Fields] OR rulemonkey[All Fields] OR rulena[All Fields]
 OR rulence[All Fields] OR ruleneg[All Fields] OR rulenko[All Fields] OR rulentnost[All Fields] OR
 rulentnost'[All Fields] OR rulenza[All Fields] OR ruleof[All Fields] OR ruleout[All Fields] OR ruleouts[All
 Fields] OR rulepidol[All Fields] OR rulequest[All Fields] OR ruler[All Fields] OR ruler'[All Fields] OR
 ruler's[All Fields] OR ruleran[All Fields] OR rulerlean[All Fields] OR rulermetr[All Fields] OR rulers[All
 Fields] OR rulers'[All Fields] OR rulership[All Fields] OR rules[All Fields] OR rules'[All Fields] OR
 rules''[All Fields] OR rulesbasemedicine[All Fields] OR ruleset[All Fields] OR rulesets[All Fields] OR
 rulesgit[All Fields] OR rulesi[All Fields] OR rulestrade[All Fields] OR rulet[All Fields] OR ruleta[All
 Fields] OR rulethe[All Fields] OR rulev[All Fields] OR ruleva[All Fields] OR rulevicius[All Fields] OR
 ruleville[All Fields] OR rulewicz[All Fields] OR ruleworld[All Fields] OR rulex[All Fields] OR ruley[All
 Fields] OR ruleye[All Fields] OR rulez[All Fields])) OR "validation score*" [All Fields] OR (derivation[All
 Fields] AND validation[All Fields]) OR (("sensitivity and specificity" [MeSH Terms] OR ("sensitivity" [All
 Fields] AND "specificity" [All Fields]) OR "sensitivity and specificity" [All Fields] OR "sensitivity" [All
 Fields]) AND ("sensitivity and specificity" [MeSH Terms] OR ("sensitivity" [All Fields] AND
 "specificity" [All Fields]) OR "sensitivity and specificity" [All Fields] OR "specificity" [All Fields]))))

N°2:

PROBAST Form

Wolff RF, Moons KGM, Riley RD, Whiting PF, Westwood M, Collins GS, et al. PROBAST: a tool to assess the risk of bias and applicability of prediction model studies. *Ann Intern Med*. 2019. doi:10.7326/M18-1376

PROBAST Form

PROBAST includes four steps.

Step	Task	When to complete
1	Specify your systematic review question(s)	Once per systematic review
2	Classify the type of prediction model evaluation	Once for each model of interest in each publication being assessed, for each relevant outcome
3	Assess risk of bias and applicability (per domain)	Once for each development and validation of each distinct prediction model in a publication
4	Overall judgment of risk of bias and applicability	Once for each development and validation of each distinct prediction model in a publication

If this is your first time using PROBAST, we strongly recommend reading the detailed explanation and elaboration (E&E) paper (27) and checking the examples on www.probast.org.

Step 1: Specify your systematic review question

State your systematic review question to facilitate the assessment of the applicability of the evaluated models to your question. <i>The following table should be completed once per systematic review.</i>	
Criteria	Specify your systematic review question
<i>Intended use of model:</i>	
Participants <i>including selection criteria and setting:</i>	
Predictors (used in modeling) <i>including (1) types of predictors (e.g. history, clinical examination, biochemical markers, imaging tests), (2) time of measurement, (3) specific measurement issues (e.g. any requirements/prohibitions for specialized equipment):</i>	
Outcome <i>to be predicted:</i>	

Step 2: Classify the type of prediction model evaluation

Use the following table to classify the evaluation as model development, model validation, or combination. Different signaling questions apply for different types of prediction model evaluation. When a publication focuses on adding one or more new predictors to established predictors then use “development only”. When a publication focuses on validation of an existing model in other data though followed by updating (adjusting or extending) of the model such that in fact a new model is being developed, then use “development and validation in the same publication”. If the evaluation does not fit one of these classifications then PROBAST should not be used.

Classify the evaluation based on its aim

Type of prediction study	PROBAST boxes to complete	Tick as appropriate	Definitions for type of prediction model study
Development only	Dev		Prediction model development without external validation. These studies may include internal validation methods such as bootstrapping and cross-validation techniques
Development and validation	Dev and Val		Prediction model development combined with external validation in other participants in the same article
Validation only	Val		External validation of existing (previously developed) model in other participants

This table should be completed once for each publication being assessed and for each relevant outcome in your review.

Publication reference	
Models of interest	
Outcome of interest	

Step 3: Assess risk of bias and applicability

PROBAST is structured as four key domains. Each domain is judged for risk of bias (low, high or unclear) and includes signaling questions to help make judgments. Signaling questions are rated as yes (Y), probably yes (PY), probably no (PN), no (N) or no information (NI). All signaling questions are phrased so that “yes” indicates absence of bias. Any signaling question rated as “no” or “probably no” flags the potential for bias; you will need to use your judgment to determine whether the domain should be rated as “high”, “low” or “unclear” risk of bias. The guidance document contains further instructions and examples on rating signaling questions and risk of bias for each domain.

The first three domains are also rated for concerns for applicability (low/ high/ unclear) to

DOMAIN 1: Participants			
A. Risk of Bias			
<i>Describe the sources of data and criteria for participant selection:</i>			
		De v	Val
1.1 Were appropriate data sources used, e.g. cohort, RCT or nested case-control study data?			
1.2 Were all inclusions and exclusions of participants appropriate?			
Risk of bias introduced by selection of participants	RISK: (low/ high/ unclear)		
<i>Rationale of bias rating:</i>			
B. Applicability			
<i>Describe included participants, setting and dates:</i>			
Concern that the included participants and setting do not match the review question	CONCERN: (low/ high/ unclear)		
<i>Rationale of applicability rating:</i>			

DOMAIN 2: Predictors			
A. Risk of Bias			
<i>List and describe predictors included in the final model, e.g. definition and timing of assessment:</i>			
		De v	Val
2.1 Were predictors defined and assessed in a similar way for all participants?			
2.2 Were predictor assessments made without knowledge of outcome data?			
2.3 Are all predictors available at the time the model is intended to be used?			
Risk of bias introduced by predictors or their assessment	RISK: (low/ high/		

	<i>unclear)</i>		
<i>Rationale of bias rating:</i>			
B. Applicabilit			
Concern that the definition, assessment or timing of predictors in the model do not match the review question	CONCERN: <i>(low/ high/ unclear)</i>		
<i>Rationale of applicability rating:</i>			

DOMAIN 3: Outcome			
A. Risk of Bias			
<i>Describe the outcome, how it was defined and determined, and the time interval between predictor assessment and outcome determination:</i>			
		Dev	Val
3.1 Was the outcome determined appropriately?			
3.2 Was a pre-specified or standard outcome definition used?			
3.3 Were predictors excluded from the outcome definition?			
3.4 Was the outcome defined and determined in a similar way for all participants?			
3.5 Was the outcome determined without knowledge of predictor information?			
3.6 Was the time interval between predictor assessment and outcome determination appropriate?			
Risk of bias introduced by the outcome or its determination	RISK: (low/ high/ unclear)		
<i>Rationale of bias rating:</i>			
B. Applicability			
<i>At what time point was the outcome determined:</i>			
<i>If a composite outcome was used, describe the relative frequency/distribution of each contributing outcome:</i>			
Concern that the outcome, its definition, timing or determination do not match the review question	CONCERN: (low/ high/ unclear)		
<i>Rationale of applicability rating:</i>			
DOMAIN 4: Analysis			
Risk of Bias			
<i>Describe numbers of participants, number of candidate predictors (for DEV only), outcome events and events per candidate predictor (for DEV only):</i>			
<i>Describe how the model was developed (predictor selection, optimism, risk groups, model performance):</i>			
<i>Describe whether and how the model was validated, either internally (e.g. bootstrapping, cross validation, random split sample) or externally (e.g. temporal validation, geographical validation, different setting, different type of participants):</i>			
<i>Describe the performance measures of the model, e.g. (re)calibration, discrimination, (re)classification, net benefit:</i>			

Describe any participants who were excluded from the analysis:			
Describe missing data on predictors and outcomes as well as methods used for missing data:			
		Dev	Val
4.1 Were there a reasonable number of participants with the outcome?			
4.2 Were continuous and categorical predictors handled appropriately?			
4.3 Were all enrolled participants included in the analysis?			
4.4 Were participants with missing data handled appropriately?			
4.5 Was selection of predictors based on univariable analysis avoided?			
4.6 Were complexities in the data (e.g. censoring, competing risks, sampling of controls) accounted for appropriately?			
4.7 Were relevant model performance measures evaluated appropriately?			
4.8 Were model overfitting and optimism in model performance accounted for?			
4.9 Do predictors and their assigned weights in the final model correspond to the results from the reported multivariable analysis?			
Risk of bias introduced by the analysis		RISK: (low/ high/ unclear)	
Rationale of bias rating:			

Step 4: Overall assessment

Use the following tables to reach overall judgments about risk of bias and concerns for applicability of the prediction model evaluation (development and/or validation) across all assessed domains.

Reaching an overall judgment about risk of bias of the prediction model evaluation	
Low risk of bias	If all domains were rated low risk of bias. If a <u>prediction model was developed without any external validation</u> , and it was rated as <u>low risk of bias for all domains</u> , consider downgrading to high risk of bias . Such a model can only be considered as low risk of bias, if the development was based on a very large data set and included some form of internal validation.
High risk of bias	If at least one domain is judged to be at high risk of bias .
Unclear risk of bias	If an unclear risk of bias was noted in at least one domain and it was low risk for all other domains.

Complete for each evaluation of a distinct model.

Reaching an overall judgment about applicability of the prediction model evaluation	
Low concerns for applicability	If low concerns for applicability for all domains, the prediction model evaluation is judged to have low concerns for applicability .
High concerns for applicability	If high concerns for applicability for at least one domain, the prediction model evaluation is judged to have high concerns for applicability .

Unclear concerns for applicability	If unclear concerns (but no “high concern”) for applicability for at least one domain, the prediction model evaluation is judged to have unclear concerns for applicability overall.	
Overall judgment about risk of bias and applicability of the prediction model evaluation		
Overall judgment of risk of bias	RISK: (low/ high/ unclear)	
<i>Summary of sources of potential bias:</i>		
Overall judgment of applicability	CONCERN: (low/ high/ unclear)	
<i>Summary of applicability concerns:</i>		

PROBAST = Prediction model Risk Of Bias ASsessment Tool.

Step 1: Specify your systematic review question

State your systematic review question to facilitate the assessment of the applicability of the evaluated models to your question. <i>The following table should be completed once per systematic review.</i>	
Criteria	Specify your systematic review question
<i>Intended use of model:</i>	<i>Review of predictive models of acute heart failure and its applicability in the hospital setting</i>
Participants including selection criteria and setting:	<i>Any original article in any language that reports one or more prediction models was included using a multivariable analysis to assess the development or outcome of acute heart failure in the adult population. In addition, it must have the description of the type of validation, discrimination report, calibration, reclassification and / or clinical utility (described as net benefit or clinical applicability) of the models developed.</i>
Predictors (used in modeling) including (1) types of predictors (e.g. history, clinical examination, biochemical markers, imaging tests), (2) time of measurement, (3) specific measurement issues (e.g. any requirements/prohibitions for specialized equipment):	<i>Any type of predictor can be used in the selected prognostic models.</i>
Outcome to be predicted:	<i>Death, exacerbation of symptoms or hospital readmission</i>
:	

--	--

Arenja N, et al: Step 2: Classify the type of prediction model evaluation

Use the following table to classify the evaluation as model development, model validation, or combination. Different signaling questions apply for different types of prediction model evaluation. When a publication focuses on adding one or more new predictors to established predictors then use “development only”. When a publication focuses on validation of an existing model in other data though followed by updating (adjusting or extending) of the model such that in fact a new model is being developed, then use “development and validation in the same publication”. If the evaluation does not fit one of these classifications then PROBAST should not be used.

Classify the evaluation based on its aim

Type of prediction study	PROBAST boxes to complete	Tick as appropriate	Definitions for type of prediction model study
Development only	Dev	X	Prediction model development without external validation. These studies may include internal validation methods such as bootstrapping and cross-validation techniques
Development and validation	Dev and Val		Prediction model development combined with external validation in other participants in the same article
Validation only	Val		External validation of existing (previously developed) model in other participants
<i>This table should be completed once for each publication being assessed and for each relevant outcome in your review.</i>			
Publication reference	Arenja, N, Breidthardt T, Socrates T, Schindler C, Heinisch C, Tschung C, et al. Risk stratification for 1-year mortality in acute heart failure. Swiss Med Wkly. 2011;141:w13259		
Models of interest	Risk stratification for 1-year mortality in acute heart failure		
Outcome of interest	Early risk stratification in-hospital mortality in patients hospitalised with AHF in a one year follow up		

Step 3: Assess risk of bias and applicability

PROBAST is structured as four key domains. Each domain is judged for risk of bias (low, high or unclear) and includes signaling questions to help make judgments. Signaling questions are rated as yes (Y), probably yes (PY), probably no (PN), no (N) or no information (NI). All signaling questions are phrased so that “yes” indicates absence of bias. Any signaling question rated as “no” or “probably no” flags the potential for bias; you will need to use your judgment to determine whether the domain should be rated as “high”, “low” or “unclear” risk of bias. The guidance document contains further instructions and examples on rating signaling questions and risk of bias for each domain. The first three domains are also rated for concerns for applicability (low/ high/ unclear) to

DOMAIN 1: Participants			
A. Risk of Bias			
<i>Describe the sources of data and criteria for participant selection:</i> Detailed data were collected from initial presentation at the ED until discharge, transfer or in-hospital death. Patients with cardiogenic shock or end-stage kidney disease (defined as serum creatinine levels of more than 250 µmol/l in the first series and by haemodialysis in the second period) were excluded, as these patients are already known to be at a high risk of death.			
		De v	Val
1.1 Were appropriate data sources used, e.g. cohort, RCT or nested case-control study data?		Y	
1.2 Were all inclusions and exclusions of participants appropriate?		Y	
Risk of bias introduced by selection of participants	RISK: (low/ high/ unclear)	low	
<i>Rationale of bias rating:</i>			
It has clear and adequate inclusion and exclusion criteria that allow to obtain a representative population			
B. Applicability			
<i>Describe included participants, setting and dates:</i> The study population consisted of consecutive patients with AHF presenting to the ED, University Hospital of Basel (Switzerland), during two periods: May 2001 to April 2002 and April 2006 to March 2007. Patient recruitment had to be paused between 2003 and 2005 due to a lack of resources.			
Concern that the included participants and setting do not match the review question	CONCERN: (low/ high/ unclear)	Low	
<i>Rationale of applicability rating:</i>			
Adequate description of patient acquisition			

DOMAIN 2: Predictors			
A. Risk of Bias			
<i>List and describe predictors included in the final model, e.g. definition and timing of assessment:</i> Of the 39 variables evaluated, the CART method identified the following three variables as predictors of mortality: BUN levels of ≥43 mg/dL (15 mmol/L), systolic blood pressure at a discrimination level of less than 115 mm Hg and a serum creatinine level of >2.75 mg/dL (243.1 µmol/L) at admission. This analysis also found BUN to be the			

strongest predictor of in-hospital mortality.			
		De v	Val
2.1 Were predictors defined and assessed in a similar way for all participants?		Y	
2.2 Were predictor assessments made without knowledge of outcome data?		Y	
2.3 Are all predictors available at the time the model is intended to be used?		Y	
Risk of bias introduced by predictors or their assessment	RISK: (low/ high/ unclear)	Low	
<i>Rationale of bias rating: The article has an adequate selection of predictors through CART, in addition to describing the cut-off points used.</i>			
B. Applicability			
Concern that the definition, assessment or timing of predictors in the model do not match the review question	CONCERN: (low/ high/ unclear)	Low	
<i>Rationale of applicability rating: They use predictors that are easy to apply and match the research question.</i>			

DOMAIN 3: Outcome			
A. Risk of Bias			
<i>Describe the outcome, how it was defined and determined, and the time interval between predictor assessment and outcome determination: We performed CART analyses, as a multivariate statistical technique, to identify the best predictors of 1-year mortality and develop a risk stratification model. The CART method allows the construction of an easily interpretable decision tree that can be applied in clinical practice. Patients were contacted for follow-up via telephone by trained researchers at 90, 180 and 360 days after discharge. Regarding mortality data, referring physicians were contacted or the administrative database of respective hometowns were reviewed if necessary. A one year follow-up was completed in 100% of patients.</i>			
		Dev	Val
3.1 Was the outcome determined appropriately?		PY	
3.2 Was a pre-specified or standard outcome definition used?		N	
3.3 Were predictors excluded from the outcome definition?		NI	
3.4 Was the outcome defined and determined in a similar way for all participants?		PY	
3.5 Was the outcome determined without knowledge of predictor information?		Y	
3.6 Was the time interval between predictor assessment and outcome determination appropriate?		PY	
Risk of bias introduced by the outcome or its determination	RISK: (low/ high/ unclear)	Unclear	
<i>Rationale of bias rating:</i> The article has a clear definition of the desired outcome, in addition to correct monitoring.			
B. Applicability			
<i>At what time point was the outcome determined: at the end of the follow-up</i>			
<i>If a composite outcome was used, describe the relative frequency/distribution of each contributing outcome:</i>			

Concern that the outcome, its definition, timing or determination do not match the review question	CONCERN: (low/ high/ unclear)	Low	
Rationale of applicability rating: 100% of the population was followed up for one year.			

DOMAIN 4: Analysis			
Risk of Bias			
Describe numbers of participants, number of candidate predictors (for DEV only), outcome events and events per candidate predictor (for DEV only): 610 patients, 38 candidate variables for predictors At 360 days, all-cause mortality was 33% (201). EPV: 5.2.			
Describe how the model was developed (predictor selection, optimism, risk groups, model performance): We performed CART analyses, as a multivariate statistical technique, to identify the best predictors of 1-year mortality and develop a risk stratification model. The CART method allows the construction of an easily interpretable decision tree that can be applied in clinical practice. As a first step, the CART model calculates parameters with cut-off points, which differentiate optimally between the study groups. These cut-offs split the data into constituent groups. In a second step, parameters are calculated, which offer the best further differentiation between the subgroups. Based on the detected predictors for 1-year mortality, patients were classified into risk groups. From these risk groups a regression tree was developed and differences regarding 1-year outcome were analysed.			
Describe whether and how the model was validated, either internally (e.g. bootstrapping, cross validation, random split sample) or externally (e.g. temporal validation, geographical validation, different setting, different type of participants): We used cross-validation to internally validate the results.			
Describe the performance measures of the model, e.g. (re)calibration, discrimination, (re)classification, net benefit: The Kaplan-Meier analysis according to the risk groups at presentation demonstrated lower one-year mortality in low risk (22%) and intermediate risk groups (37%) compared to the high risk group (63%) (Log-rank test $p < 0.001$; figure 3). The HR regarding 1-year mortality between the high and low risk groups was 1.9 (95% CI, 1.7–2.2), between the high and intermediate risk groups it was 2 (95% CI, 1.5–2.7), and between the intermediate and low risk groups it was 1.7 (95% CI, 1.3–2.3), with statistically significant differences between all risk groups ($p < 0.001$; table 4). The likelihood-based 95%-confidence set for the age- and the urea-threshold is contained in the rectangular set defined by $25 \text{ mg/dl} \leq \text{urea threshold} \leq 30.6 \text{ mg/dl}$ and $76 \text{ years} \leq \text{age threshold} \leq 96 \text{ years}$.			
Describe any participants who were excluded from the analysis: Patients with cardiogenic shock or end-stage kidney disease (defined as serum creatinine levels of more than $250 \mu\text{mol/l}$ in the first series and by haemodialysis in the second period) were excluded, as these patients are already known to be at a high risk of death			
Describe missing data on predictors and outcomes as well as methods used for missing data: The article mentions that 100% of the population was monitored.			
		Dev	Val
4.1 Were there a reasonable number of participants with the outcome?		N	
4.2 Were continuous and categorical predictors handled appropriately?		Y	
4.3 Were all enrolled participants included in the analysis?		Y	
4.4 Were participants with missing data handled appropriately?		Y	
4.5 Was selection of predictors based on univariable analysis avoided?		PY	
4.6 Were complexities in the data (e.g. censoring, competing risks, sampling of controls) accounted for appropriately?		NI	
4.7 Were relevant model performance measures evaluated appropriately?		N	
4.8 Were model overfitting and optimism in model performance accounted for?		N	
4.9 Do predictors and their assigned weights in the final model correspond to the results from the reported multivariable analysis?		PY	
Risk of bias introduced by the analysis	RISK: (low/ unclear)	high/	High

Rationale of bias rating: Describe adequately the selection of predictors and the performance of the scale.

Step 4: Overall assessment

Use the following tables to reach overall judgments about risk of bias and concerns for applicability of the prediction model evaluation (development and/or validation) across all assessed domains.

Reaching an overall judgment about risk of bias of the prediction model evaluation	
Low risk of bias	If all domains were rated low risk of bias. If a prediction model was developed without any external validation, and it was rated as <u>low risk of bias for all domains</u> , consider downgrading to high risk of bias . Such a model can only be considered as low risk of bias, if the development was based on a very large data set and included some form of internal validation.
High risk of bias	If at least one domain is judged to be at high risk of bias .
Unclear risk of bias	If an unclear risk of bias was noted in at least one domain and it was low risk for all other domains.

Complete for each evaluation of a distinct model.

Reaching an overall judgment about applicability of the prediction model evaluation	
Low concerns for applicability	If low concerns for applicability for all domains, the prediction model evaluation is judged to have low concerns for applicability .
High concerns for applicability	If high concerns for applicability for at least one domain, the prediction model evaluation is judged to have high concerns for applicability .
Unclear concerns for applicability	If unclear concerns (but no “high concern”) for applicability for at least one domain, the prediction model evaluation is judged to have unclear concerns for applicability overall.

Overall judgment about risk of bias and applicability of the prediction model evaluation		
Overall judgment of risk of bias	RISK: (low/ high/ unclear)	High
<i>Summary of sources of potential bias: High</i>		
Overall judgment of applicability	CONCERN: (low/ high/ unclear)	low
<i>Summary of applicability concerns: Low</i>		

PROBAST = Prediction model Risk Of Bias ASsessment Tool.

1. Behnes M, et al

Step 2: Classify the type of prediction model evaluation

Classify the evaluation based on its aim			
Type of prediction study	PROBAST boxes to complete	Tick as appropriate	Definitions for type of prediction model study
Development only	Dev	X	Prediction model development without external validation. These studies may include internal validation methods such as bootstrapping and cross-validation techniques
Development and validation	Dev and Val		Prediction model development combined with external validation in other participants in the same article
Validation only	Val		External validation of existing (previously developed) model in other participants
<i>This table should be completed once for each publication being assessed and for each relevant outcome in your review.</i>			
Publication reference	Behnes M, Bertsch T, Weiss C, Ahmad-Nejad P, Akin I, Fastner C, El-Battrawy I, et al. Triple head-to-head comparison of fibrotic biomarkers galectin-3, osteopontin and gremlin-1 for long-term prognosis in suspected and proven acute heart failure patients. International Journal of Cardiology 203 (2016) 398–406.		
Models of interest	Triple head-to-head comparison of fibrotic biomarkers galectin-3, osteopontin and gremlin-1 for long-term prognosis in suspected and proven acute heart failure patients		
Outcome of interest	Prognostic outcomes were long-term all-cause mortality (ACM) and AHF-related rehospitalization (AHF-RH) at 1 and 5 years.		

Step 3: Assess risk of bias and applicability

DOMAIN 1: Participants		
A. Risk of Bias		
<i>Describe the sources of data and criteria for participant selection:</i> The present study represents a post hoc analysis of a specimen repository from patients enrolled in the Mannheim NT-proBNP Study (MANPRO), which was conducted as a mono-centric prospective controlled study at the University Medical Centre Mannheim (UMM), Germany. Inclusion criteria of the MANPRO study were patients with symptoms of acute dyspnea and/or peripheral edema presenting in the ED. Patients suffering from severe renal disease (defined as serum creatinine level greater than 2.8 mg/dl), anemia (hemoglobin concentrations below 8.0 g/dl), obvious traumatic causes of dyspnea, pregnancy, with a status after immediate cardiopulmonary resuscitation, participation in another clinical trial and patients with age under 18 years were excluded.		
	Dev	Val
1.1 Were appropriate data sources used, e.g. cohort, RCT or nested case-control study data?	Y	

1.2 Were all inclusions and exclusions of participants appropriate?		Y	
Risk of bias introduced by selection of participants	RISK: (low/unclear) high/	low	
<i>Rationale of bias rating: It has a database with appropriate selection and exclusion criteria.</i>			
B. Applicability			
<i>Describe included participants, setting and dates:</i> Patients, which were consecutively included from August 2005 until March 2006. Baseline characteristics of the whole study population (n=401) as well as the subgroup of AHF patients (n = 122) In AHF patients, 75% (n= 92) suffered from heart failure with reduced ejection fraction (LVEF b55%), whereas 25% (n= 30) suffered from heart failure with preserved ejection fraction (LVEF N55%).			
Concern that the included participants and setting do not match the review question	CONCERN: (low/unclear) high/	Low	
<i>Rationale of applicability rating: the selection of participants was appropriate to the population with AHF</i>			

DOMAIN 2: Predictors			
A. Risk of Bias			
<i>List and describe predictors included in the final model, e.g. definition and timing of assessment:</i> age, sex, left ventricular function, serum creatinine (mg/dl), NYHA functional class, presence of diabetes mellitus, the degree of coronary artery disease, hemoglobin (g/dl), serum sodium (mmol/l), beta-blocker treatment, angiotensin-converting enzyme inhibitor or angiotensin II receptor blocker treatment (ACEI/ARB) and logNT-proBNP levels. Gremlin-1, osteopontin and galectin-3 were subsequently added to this reference model. All samples were obtained by venipuncture into serum and ammonium heparin tubes for biomarker measurements, immediately at presentation to the ED. Within 30 min all blood samples were centrifuged at 2000 g for 10 min. Plasma and serum were separated, aliquoted, frozen and stored at -80 °C. Blood samples for biomarker measurements were available in all patients. Biomarker measurement was performed by commercially available enzyme-linked immunosorbent assays: osteopontin (Quantikine Human Osteopontin (OPN) Immunoassay (R&D Systems Inc., Minneapolis, USA)), gremlin-1 (Human Grem1 ELISA®, USC Life Science Inc., Wuhan, China), galectin-3 (Galectin 3, Architect System, Abbott, Wiesbaden, Germany) and NT-proBNP (Flex reagent cartridge PBNP, Dimension System, Dade Behring Ltd., Atterbury, Milton Keynes, UK) as previously described.			
		De v	Val
2.1 Were predictors defined and assessed in a similar way for all participants?		Y	
2.2 Were predictor assessments made without knowledge of outcome data?		Y	
2.3 Are all predictors available at the time the model is intended to be used?		Y	
Risk of bias introduced by predictors or their assessment	RISK: (low/unclear) high/	Low	
<i>Rationale of bias rating: They describe the variables included in each predictive model, which were evaluated during the follow-up, at the appropriate time. They also indicate the reference of which they were chosen</i>			
B. Applicability			
Concern that the definition, assessment or timing of predictors in the model do not match the review question	CONCERN: (low/unclear) high/	Low	
<i>Rationale of applicability rating: Biomarkers can be applied in a correct time suitable for risk classification.</i>			
DOMAIN 3: Outcome			
A. Risk of Bias			

<i>Describe the outcome, how it was defined and determined, and the time interval between predictor assessment and outcome determination:</i> Two prognostic outcomes were considered: ACM and AHF-RH at 1 and 5 years. Follow-ups were performed in three successive steps using our in-hospital electronic records, contacting family physicians and performing individual telephone visits. The present study complements and expands the spectrum of scientific evidence with regard to long-term prognostic implications of galectin-3, osteopontin and gremlin-1 in patients suspected to suffer from AHF over a follow-up period of up to 5 years.			
		Dev	Val
3.1 Was the outcome determined appropriately?		Y	
3.2 Was a pre-specified or standard outcome definition used?		Y	
3.3 Were predictors excluded from the outcome definition?		Y	
3.4 Was the outcome defined and determined in a similar way for all participants?		Y	
3.5 Was the outcome determined without knowledge of predictor information?		Y	
3.6 Was the time interval between predictor assessment and outcome determination appropriate?		Y	
Risk of bias introduced by the outcome or its determination	RISK: (low/ high/ unclear)	Low	
<i>Rationale of bias rating:</i> The model provides a clear definition of the outcome that was sought, in addition to correct monitoring, from admission to emergency until the presentation of the event.			
B. Applicability			
<i>At what time point was the outcome determined:</i> <i>It was determined during the follow-up in the future.</i>			
<i>If a composite outcome was used, describe the relative frequency/distribution of each contributing outcome:</i>			
Concern that the outcome, its definition, timing or determination do not match the review question	CONCERN: (low/ high/ unclear)	Low	
<i>Rationale of applicability rating:</i> The time in which the outcome is determined is consistent with the proposed research question.			

DOMAIN 4: Analysis			
Risk of Bias			
Describe numbers of participants, number of candidate predictors (for DEV only), outcome events and events per candidate predictor (for DEV only): The population included in the study is 401. Candidate predictors are not reported, so the EPV cannot be estimated			
Describe how the model was developed (predictor selection, optimism, risk groups, model performance): Prognostic models were performed using Cox regression analyses including the following 12 variables to a reference model: age, sex, left ventricular function, serum creatinine (mg/dl), NYHA functional class, presence of diabetes mellitus, the degree of coronary artery disease, hemoglobin (g/dl), serum sodium (mmol/l), beta-blocker treatment, angiotensin-converting enzyme inhibitor or angiotensin II receptor blocker treatment (ACEI/ARB) and logNT-proBNP levels. Gremlin-1, osteopontin and galectin-3 were subsequently added to this reference model. Hazard Ratios (HR) were calculated for each risk group according to quartiles of each biomarker. The long-term prognostic value was tested in prognostic models including 11 conventional risk factors plus NT-proBNP and was assessed by state-of-the-art statistics of discrimination, calibration, and reclassification and Cox regression analyses.			
Describe whether and how the model was validated, either internally (e.g. bootstrapping, cross validation, random split sample) or externally (e.g. temporal validation, geographical validation, different setting, different type of participants): not mentioned			
Describe the performance measures of the model, e.g. (re)calibration, discrimination, (re)classification, net benefit: Discrimination: All three fibrotic biomarkers were adjusted in multivariable prognostic models relative to conventional assessment (i.e. 11 risk factors plus NT-pro BNP). All multivariable prognostic models discriminated individually the prognostic endpoints of ACM and AHF-RH even after adding each fibrotic biomarker either to all patients or to the subgroup of AHF patients (p=0.0001, referring to each model). Calibration and reclassification ACM in all patients was best calibrated in prognostic models including galectin-3. These models revealed favorable measures of calibration (such as Hosmer–Lemeshow, Brier Score, AIC, BIC and likelihood ratios) as well as beneficial NRIs and IDIs (Table 3A, top, highlighted in blue). ACM in AHF patients was best calibrated in models using osteopontin, which accordingly resulted in beneficial reclassification of these patients (Table 3A, bottom, highli). AHF-RH at 1 (Tabl)and 5 years (data not shown) was best calibrated in prognostic models using gremlin-1 both in all and AHF patientsand accordingly resulted in beneficial reclassification of these patients.			
Describe any participants who were excluded from the analysis: Patients suffering from severe renal disease (defined as serum creatinine level greater than 2.8 mg/dl), anemia (hemoglobin concentrations below 8.0 g/dl), obvious traumatic causes of dyspnea, pregnancy, with a status after immediate cardiopulmonary resuscitation, participation in another clinical trial and patients with age under 18 years were excluded			
Describe missing data on predictors and outcomes as well as methods used for missing data: not mentioned			
		Dev	Val
4.1 Were there a reasonable number of participants with the outcome?		Y	
4.2 Were continuous and categorical predictors handled appropriately?		Y	
4.3 Were all enrolled participants included in the analysis?		Y	
4.4 Were participants with missing data handled appropriately?		NI	
4.5 Was selection of predictors based on univariable analysis avoided?		Y	
4.6 Were complexities in the data (e.g. censoring, competing risks, sampling of controls) accounted for appropriately?		NI	
4.7 Were relevant model performance measures evaluated appropriately?		Y	
4.8 Were model overfitting and optimism in model performance accounted for?		Y	
4.9 Do predictors and their assigned weights in the final model correspond to the results from the reported multivariable analysis?		Y	
Risk of bias introduced by the analysis	RISK: (low/ unclear)	Low	

Rationale of bias rating:

Even with the missing data regarding lost information, the model adequately explains the performance, so it is considered low risk of bias.

Step 4: Overall assessment

Use the following tables to reach overall judgments about risk of bias and concerns for applicability of the prediction model evaluation (development and/or validation) across all assessed domains.

Reaching an overall judgment about risk of bias of the prediction model evaluation

Low risk of bias	If all domains were rated low risk of bias. If a <u>prediction model was developed without any external validation</u> , and it was rated as <u>low risk of bias for all domains</u> , consider downgrading to high risk of bias . Such a model can only be considered as low risk of bias, if the development was based on a very large data set and included some form of internal validation.
High risk of bias	If at least one domain is judged to be at high risk of bias .
Unclear risk of bias	If an unclear risk of bias was noted in at least one domain and it was low risk for all other domains.

Complete for each evaluation of a distinct model.

Reaching an overall judgment about applicability of the prediction model evaluation

Low concerns for applicability	If low concerns for applicability for all domains, the prediction model evaluation is judged to have low concerns for applicability .
High concerns for applicability	If high concerns for applicability for at least one domain, the prediction model evaluation is judged to have high concerns for applicability .
Unclear concerns for applicability	If unclear concerns (but no “high concern”) for applicability for at least one domain, the prediction model evaluation is judged to have unclear concerns for applicability overall.

Overall judgment about risk of bias and applicability of the prediction model evaluation

Overall judgment of risk of bias	RISK: (low/ high/ unclear)	
<i>Summary of sources of potential bias: Low</i>		
Overall judgment of applicability	CONCERN: (low/ high/ unclear)	

Summary of applicability concerns: Low

PROBAST = Prediction model Risk Of Bias ASsessment Tool.

2. Borolkar KA, et al.

Step 2: Classify the type of prediction model evaluation

Classify the evaluation based on its aim			
Type of prediction study	PROBAST boxes to complete	Tick as appropriate	Definitions for type of prediction model study
Development only	Dev	X	Prediction model development without external validation. These studies may include internal validation methods such as bootstrapping and cross-validation techniques
Development and validation	Dev and Val		Prediction model development combined with external validation in other participants in the same article
Validation only	Val		External validation of existing (previously developed) model in other participants
<i>This table should be completed once for each publication being assessed and for each relevant outcome in your review.</i>			
Publication reference	Borolkar KA, Kobayashi Y, Moneghetti KJ, Pargaonkar VS, Tuzovic M, Krishnan G, et al. Improving risk stratification in heart failure with preserved ejection fraction by combining two validated risk scores. Open Heart 2019;6:e000961. doi:10.1136/openhrt-2018-000961		
Models of interest	IMRS and GWTG-HF risk scores. Plus NT-proBNP.		
Outcome of interest	Hypothesised that the IMRS would predict all-cause mortality specifically in patients hospitalised with HFpEF; in addition, we further hypothesised that it would be complementary to the GWTG-HF risk score or N-terminal pro B-type natriuretic peptide (NT-proBNP). We also sought to better delineate the relationship between variables using correlation network analysis.		

Step 3: Assess risk of bias and applicability

DOMAIN 1: Participants
A. Risk of Bias
<i>Describe the sources of data and criteria for participant selection:</i> We used the Stanford Translational Research Integrated Database Environment to identify 580 adult patients with a diagnosis of HFpEF between January 1998 and December 2016. Of the identified patients, we only selected patients who underwent an echocardiogram within the year of hospitalisation to verify the accurate evaluation of ejection fraction. Of these patients, 1923 had an admission for acute HF. We excluded patients with left ventricular ejection fraction (LVEF) <50% (n=288), hypertrophic cardiomyopathy (n=129), pulmonary arterial hypertension (n=41), heart

transplantation (n=62), severe valvular heart disease (n=348), pericardial heart disease (n=5), end-stage renal disease and advanced liver disease (n=280) or active cancer (n=190). Finally, 580 patients met the inclusion and exclusion criteria and were included in the analysis			
		De v	Val
1.1 Were appropriate data sources used, e.g. cohort, RCT or nested case-control study data?		Y	
1.2 Were all inclusions and exclusions of participants appropriate?		Y	
Risk of bias introduced by selection of participants	RISK: (low/ high/ unclear)	Low	
<i>Rationale of bias rating:</i>			
Not so clear the inclusion criteria but well explained the exclusion criteria.			
B. Applicability			
Describe included participants, setting and dates: adult patients with a diagnostic code for acute HFpEF between January 1998 and December 2016. The mean age of the population was 76±16 years with 334 (58%) female and a significant proportion of patients with DM (36%), a history of systemic hypertension (97%) or a previous history of Afib (58%). The median duration of hospitalisation was 4 (IQR 2–9) days.			
Concern that the included participants and setting do not match the review question	CONCERN: (low/ high/ unclear)	Low	
<i>Rationale of applicability rating: Information match the review question.</i>			

DOMAIN 2: Predictors			
A. Risk of Bias			
<i>List and describe predictors included in the final model, e.g. definition and timing of assessment: Predictors of GWTG-HF, IMRS and NT-proBNP. Age, sex, race, comorbidities, smoking, sbp, dbp, HR, RR, O2, BMI, WBC count, Hb, Hct, Platelets, MPV, RDW, MCV, MCH, MCHC, hsCRP, HDL, LDL, Sodium, BUN, Cr, Potassium, Chloride, Bicarbonate, Calcium, Glucose, LVEF, plus NT-proBNP</i>			
		De v	Val
2.1 Were predictors defined and assessed in a similar way for all participants?		Y	
2.2 Were predictor assessments made without knowledge of outcome data?		Y	
2.3 Are all predictors available at the time the model is intended to be used?		Y	
Risk of bias introduced by predictors or their assessment	RISK: (low/ high/ unclear)	Low	
<i>Rationale of bias rating: The article has a good selection of variables combining variables known as risk factors in AHF in 2 different scores.</i>			
B. Applicability			
Concern that the definition, assessment or timing of predictors in the model do not match the review question	CONCERN: (low/ high/ unclear)	Uncle ar	
<i>Rationale of applicability rating: They use 2 scores, combining them and adding one variable, so maybe the applicability is not so easy,</i>			
DOMAIN 3: Outcome			
A. Risk of Bias			

Describe the outcome, how it was defined and determined, and the time interval between predictor assessment and outcome determination: Hypothesised that the IMRS would predict all-cause mortality specifically in patients hospitalised with HFpEF; in addition, we further hypothesised that it would be complementary to the GWTG-HF risk score or N-terminal pro B-type natriuretic peptide (NT-proBNP). They do not specified time of prediction.			
		Dev	Val
3.1 Was the outcome determined appropriately?		Y	
3.2 Was a pre-specified or standard outcome definition used?		Y	
3.3 Were predictors excluded from the outcome definition?		Y	
3.4 Was the outcome defined and determined in a similar way for all participants?		Y	
3.5 Was the outcome determined without knowledge of predictor information?		Y	
3.6 Was the time interval between predictor assessment and outcome determination appropriate?		Y	
Risk of bias introduced by the outcome or its determination	RISK: (low/ high/ unclear)	Low	
Rationale of bias rating: It was very well defined outcome, with 2 good models			
B. Applicability			
At what time point was the outcome determined: at one time in the future			
If a composite outcome was used, describe the relative frequency/distribution of each contributing outcome:			
Concern that the outcome, its definition, timing or determination do not match the review question	CONCERN: (low/ high/ unclear)	Low	
Rationale of applicability rating: Outcome fit with the review question			
DOMAIN 4: Analysis			
Risk of Bias			
Describe numbers of participants, number of candidate predictors (for DEV only), outcome events and events per candidate predictor (for DEV only): 580 patients met the inclusion and exclusion criteria and were included in the analysis			
Describe how the model was developed (predictor selection, optimism, risk groups, model performance): On multivariable Cox proportional hazard analysis, IMRS, GWTG-HF risk score and log NT-proBNP were independently associated with overall all-cause mortality with a log likelihood ratio of -573.90. When the scores were excluded from the multivariable model, age, heart rate, MCV, MCHC, RDW, sodium and BUN and log NT-proBNP were independently associated with overall all-cause mortality with a log likelihood ratio of -562.80. On multivariable Cox proportional hazard analysis, IMRS, GWTG-HF risk score and log NT-proBNP were inde- pendently associated with overall all-cause mortality with a log likelihood ratio of -573.90. When the scores were excluded from the multivariable model, age, heart rate, MCV, MCHC, RDW, sodium and BUN and log NT-proBNP were independently associated with overall all-cause mortality with a log likelihood ratio of -562.80			

Describe whether and how the model was validated, either internally (e.g. bootstrapping, cross validation, random split sample) or externally (e.g. temporal validation, geographical validation, different setting, different type of participants): They do not validated this score. In the limitations, they mention validation is required. They develop 1 score combining 2 previous scores and NT-proBNP			
Describe the performance measures of the model, e.g. (re)calibration, discrimination, (re)classification, net benefit: NT-proBNP improved the net reclassification of both scores. BNP or NT-proBNP have been used as a supportive diagnostic criteria for HFpEF. Combining the two scores, improved the net reclassification over GWTG-HF alone by 36.2%. In patients with available NT-proBNP (n=341), NT-proBNP improved the net reclassification of each score by 46.2% (IMRS) and 36.3% (GWTG-HF).			
Describe any participants who were excluded from the analysis: patients with left ventricular ejection fraction (LVEF) <50%, hypertrophic cardiomyopathy, pulmonary arterial hypertension, heart transplantation, severe valvular heart disease, pericardial heart disease, end-stage renal disease and advanced liver disease or active cancer.			
Describe missing data on predictors and outcomes as well as methods used for missing data: Do not present any missing data			
		Dev	Val
4.1 Were there a reasonable number of participants with the outcome?		PN	
4.2 Were continuous and categorical predictors handled appropriately?		Y	
4.3 Were all enrolled participants included in the analysis?		Y	
4.4 Were participants with missing data handled appropriately?		NI	
4.5 Was selection of predictors based on univariable analysis avoided?		Y	
4.6 Were complexities in the data (e.g. censoring, competing risks, sampling of controls) accounted for appropriately?		NI	
4.7 Were relevant model performance measures evaluated appropriately?		Y	
4.8 Were model overfitting and optimism in model performance accounted for?		Y	
4.9 Do predictors and their assigned weights in the final model correspond to the results from the reported multivariable analysis?		Y	
Risk of bias introduced by the analysis	RISK: (low/ high/ unclear)	Low	
Rationale of bias rating: Describe how they get the score			

Step 4: Overall assessment

Use the following tables to reach overall judgments about risk of bias and concerns for applicability of the prediction model evaluation (development and/or validation) across all assessed domains.

Reaching an overall judgment about risk of bias of the prediction model evaluation	
Low risk of bias	If all domains were rated low risk of bias. If a prediction model was developed without any external validation, and it was rated as <u>low risk of bias for all domains</u> , consider downgrading to high risk of bias . Such a model can only be considered as low risk of bias, if the development was based on a very large data set and included some form of internal validation.
High risk of bias	If at least one domain is judged to be at high risk of bias .
Unclear risk of bias	If an unclear risk of bias was noted in at least one domain and it was low risk for all other domains.

Complete for each evaluation of a distinct model.

Reaching an overall judgment about applicability of the prediction model evaluation	
Low concerns for applicability	If low concerns for applicability for all domains, the prediction model evaluation is judged to have low concerns for applicability .
High concerns for applicability	If high concerns for applicability for at least one domain, the prediction model evaluation is judged to have high concerns for applicability .
Unclear concerns for applicability	If unclear concerns (but no “high concern”) for applicability for at least one domain, the prediction model evaluation is judged to have unclear concerns for applicability overall.

Overall judgment about risk of bias and applicability of the prediction model evaluation		
Overall judgment of risk of bias	RISK: (low/ high/ unclear)	Low
<i>Summary of sources of potential bias: Low</i>		
Overall judgment of applicability	CONCERN: (low/ high/ unclear)	High
<i>Summary of applicability concerns: High</i>		

PROBAST = Prediction model Risk Of Bias ASsessment Tool.

3. Borovac JA,et al:

Step 2: Classify the type of prediction model evaluation

Classify the evaluation based on its aim			
Type of prediction study	PROBAST boxes to complete	Tick as appropriate	Definitions for type of prediction model study
Development only	Dev	X	Prediction model development without external validation. These studies may include internal validation methods such as bootstrapping and cross-validation techniques
Development and validation	Dev and Val		Prediction model development combined with external validation in other participants in the same article
Validation only	Val		External validation of existing (previously developed) model in other participants
<i>This table should be completed once for each publication being assessed and for each relevant outcome in your review.</i>			
Publication reference	Borovac JA, Glavas D, Bozic J, Novak K, et al. Predicting the 1-Year All-Cause Mortality After Hospitalization for an Acute Heart Failure Event: A Real-World Derivation Cohort for the Development of the S2PLiT-UG Score. Heart, Lung and Circulation (2019), https://doi.org/10.1016/j.hlc.2019.03.021		
Models of interest	S2PLiT-UG Score		
Outcome of interest	Identify independent and robust predictors of all-cause mortality during the 1-year FU period in a sample of real-world AHF patients with substantial comorbidity burden; and to characterize and develop a simple risk score that could reliably predict 1-year all-cause mortality events in this population, which in turn might help to better identify AHF patients who would likely benefit from more stringent post-discharge monitoring and/or adjunctive therapeutic strategies.		

Step 3: Assess risk of bias and applicability

DOMAIN 1: Participants		
A. Risk of Bias		
Describe the sources of data and criteria for participant selection: Data were retrospectively retrieved from a dedicated HF registry at a tertiary-care university hospital centre. This database contained baseline demographic, clinical, comorbidity, echocardiographic, and laboratory data acquired during the index admission for AHF events during February 2010 to January 2013.		
	Dev	Val

1.1 Were appropriate data sources used, e.g. cohort, RCT or nested case-control study data?	Y	
1.2 Were all inclusions and exclusions of participants appropriate?	y	
Risk of bias introduced by selection of participants	RISK: (low/ high/ unclear)	Low
<i>Rationale of bias rating:</i>		
Inclusion and Exclusion information were mentioned.		
B. Applicability		
<i>Describe included participants, setting and dates:</i> The study population consisted of patients with AHF. Total number of patients 300 (100) Women 163 (54.3) Age (years) 75.2 ± 6.6, Crude all-cause mortality rate during FU 103 (34.3), HF-related readmission events during FU 85, Ischaemic cardiomyopathy 238 (79.3), Non-ischaemic cardiomyopathy 62 (20.7), Arterial hypertension 172 (57.3), Atrial fibrillation 153 (51.0), Diabetes mellitus 80 (26.7), Chronic kidney disease 196 (65.3), Peripheral congestion 125 (41.7), Systolic blood pressure (mmHg) 140.7 ± 4.1, Diastolic blood pressure (mmHg) 82.9 ± 14.2, LVEF (%) 43.9 ± 8.7, LVEDd (mm) 62.5 ± 9.58, HFrEF 78 (26.0), HFmrEF 134 (44.7), HFpEF 88 (29.3), NYHA II 77 (25.7), NYHA III 145 (48.3), NYHA IV 78 (26.0), No previous HF-related hospitalisations 102 (34.0), One or more HF-related hospitalisations 198 (66.0), Creatinine (mmol/L) 131.2 ± 88.9, eGFR CKD-EPI (mL/min/1.73 m ²) 50.6 ± 20.6, Urea (mmol/L) 8.7 ± 6.8, Urea: Creatinine Ratio (UCR, X:1) 67.8:1 ± 37.9, Uric acid (mmol/L) 469.6 ± 163.6, Sodium (mmol/L) 138.8 ± 4.1, Potassium (mmol/L) 4.4 ± 0.7, Haemoglobin (g/L) 124.3 ± 20.2, Medication intake ACE-I/ARBs 230 (76.7), Beta-blockers 213 (71.1), Diuretics 292 (97.3)		
Concern that the included participants and setting do not match the review question	CONCERN: (low/ high/ unclear)	Low
<i>Rationale of applicability rating: Information is complete.</i>		

DOMAIN 2: Predictors			
A. Risk of Bias			
<i>List and describe predictors included in the final model, e.g. definition and timing of assessment: Systolic blood pressure <115 mmHg, Plasma sodium <136 mmol/L, LVEF <45%, Previous history of decompensation event(s) that required admission to hospital, Uric acid >450 mmol/L, eGFR CKD-EPI <60</i>			
		De v	Val
2.1 Were predictors defined and assessed in a similar way for all participants?		Y	
2.2 Were predictor assessments made without knowledge of outcome data?		Y	
2.3 Are all predictors available at the time the model is intended to be used?		Y	
Risk of bias introduced by predictors or their assessment	RISK: (low/ high/ unclear)	Low	
<i>Rationale of bias rating: The article has a good selection of variables combining variables known as risk factors in AHF.</i>			
B. Applicability			
Concern that the definition, assessment or timing of predictors in the model do not match the review question	CONCERN: (low/ high/ unclear)	Low	
<i>Rationale of applicability rating: They use a simple group of variables, easy to get for develop one usefull and very applicable model.</i>			
DOMAIN 3: Outcome			

A. Risk of Bias			
Describe the outcome, how it was defined and determined, and the time interval between predictor assessment and outcome determination: Identify independent and robust predictors of all-cause mortality during the 1-year FU period in a sample of real-world AHF patients with substantial comorbidity burden; and to characterize and develop a simple risk score that could reliably predict 1-year all-cause mortality events in this population, which in turn might help to better identify AHF patients who would likely benefit from more stringent post-discharge monitoring and/or adjunctive therapeutic strategies.			
		Dev	Val
3.1 Was the outcome determined appropriately?		Y	
3.2 Was a pre-specified or standard outcome definition used?		Y	
3.3 Were predictors excluded from the outcome definition?		Y	
3.4 Was the outcome defined and determined in a similar way for all participants?		Y	
3.5 Was the outcome determined without knowledge of predictor information?		Y	
3.6 Was the time interval between predictor assessment and outcome determination appropriate?		Y	
Risk of bias introduced by the outcome or its determination		RISK: (low/ high/ unclear)	Low
Rationale of bias rating: It was very well defined outcome, with good model			
B. Applicability			
At what time point was the outcome determined: at one time in the future			
If a composite outcome was used, describe the relative frequency/distribution of each contributing outcome:			
Concern that the outcome, its definition, timing or determination do not match the review question		CONCERN: (low/ high/ unclear)	Low
Rationale of applicability rating: It was correctly presented all the information.			
DOMAIN 4: Analysis			
Risk of Bias			
Describe numbers of participants, number of candidate predictors (for DEV only), outcome events and events per candidate predictor (for DEV only): 300 patients fulfilling inclusion criteria were identified in the database. No information for EPV.			
Describe how the model was developed (predictor selection, optimism, risk groups, model performance): Data were presented as numbers (%) for categorical and mean (M) ± standard deviation (SD) for continuous variables. The obtained data were used to identify predictors of all-cause 1-year mortality following an AHF event and hospital discharge. Furthermore, identified risk factors were used to develop a risk scoring system. A univariate logistic regression analysis was performed to determine the association between different single variables set as independent predictors, and all-cause death during the 1-year FU set as a dependent outcome in the derivation cohort. All analysed cases were complete with no missing data. Independent variables with a p-value <0.05 in the univariate analysis were entered in the Cox regression with the forward-conditional method and bootstrapping procedure with 1000 samples to derive adjusted hazard ratios (HRs) and 95% confidence intervals (95% CI). The respective regression model was adjusted for following covariates: age, sex, eGFR, number of acute decompensation events requiring hospitalization prior to current admission, NYHA functional class, diabetes mellitus, SBP, LVEF and laboratory parameters. The Hosmer-Lemeshow test was used to inspect the goodness-of-fit of the regression model with p-value >0.05 being classified as a good fit. The S2PLiT-UG scoring system was then developed by scaling and rounding significant regression coefficients obtained through a Cox regression model. The acronym S2PLiT-UG was			

<p>constructed based on the first letter of the components that constitute the scoring system: S2 = systolic blood pressure and sodium; P = plasma creatinine; L = left-ventricular ejection fraction; T = trouble before (as if the patient has a positive history of HF-related hospitalisations); U = uric acid; and G = glomerular filtration rate. Calibration of the derivation cohort was assessed by Hosmer- Lemeshow goodness-of-fit test and by plotting observed vs predicted incidence rates of all-cause mortality, and R2 value was reported for those. A receiver-operating characteristic (ROC) curve analysis with reported c-statistic (area under the curve, AUC), standard error (SE), 95% CI, and p-value was performed in order to determine the cut-off values (for which Youden index was used) and to evaluate prognostic performance of the S2PLiT-UG score. The cumulative survival and statistical difference in survival among risk-assigned groups were determined by Kaplan- Meier with log-rank (Mantel-Cox) analysis. A two-sided p value <0.05 was considered statistically significant at all instances. Data analyses were performed by using SPSS Statistics for Windows, version 25.0 (IBM Corp., Armonk, NY, USA) and Prism 6.01 for Windows (GraphPad, La Jolla, CA, USA).</p>			
<p>Describe whether and how the model was validated, either internally (e.g. bootstrapping, cross validation, random split sample) or externally (e.g. temporal validation, geographical validation, different setting, different type of participants): They develop the score by internal validation.</p>			
<p>Describe the performance measures of the model, e.g. (re)calibration, discrimination, (re)classification, net benefit: The obtained AUC value for the proposed S2PLiT-UG score in predicting 1-year all-cause mortality was 0.907 (SE 0.0183; 95% CI, 0.867-0.939; p < 0.0001) (Figure 4). Moreover, a score provided good calibration in terms of observed vs predicted incidence of all-cause death when both were plotted against each other (R2 value 0.984, Hosmer-Lemeshow goodness-of-fit p = 0.140).</p>			
<p>Describe any participants who were excluded from the analysis: NYHA I class, those who suffered in-hospital death, and those who presented with some of the following characteristics during index hospitalisation: clinically and laboratory verified active infectious or systemic autoimmune disease, immunocompromised state, cancer of any origin, significant liver disease or cirrhosis, haemorrhagic diathesis, alcohol or substance abuse, stroke or acute coronary syndrome (ACS), or major non-cardiac or cardiac surgery within 3 months prior to index admission.</p>			
<p>Describe missing data on predictors and outcomes as well as methods used for missing data: All analysed cases were complete with no missing data.</p>			
		Dev	Val
4.1 Were there a reasonable number of participants with the outcome?		Y	
4.2 Were continuous and categorical predictors handled appropriately?		Y	
4.3 Were all enrolled participants included in the analysis?		Y	
4.4 Were participants with missing data handled appropriately?		NI	
4.5 Was selection of predictors based on univariable analysis avoided?		Y	
4.6 Were complexities in the data (e.g. censoring, competing risks, sampling of controls) accounted for appropriately?		NI	
4.7 Were relevant model performance measures evaluated appropriately?		Y	
4.8 Were model overfitting and optimism in model performance accounted for?		Y	
4.9 Do predictors and their assigned weights in the final model correspond to the results from the reported multivariable analysis?		Y	
Risk of bias introduced by the analysis	RISK: (low/ high/ unclear)	Low	
<p>Rationale of bias rating: Describe how they get the score</p>			

Step 4: Overall assessment

Use the following tables to reach overall judgments about risk of bias and concerns for applicability of the prediction model evaluation (development and/or validation) across all assessed domains.

Reaching an overall judgment about risk of bias of the prediction model evaluation

Low risk of bias	If all domains were rated low risk of bias. If a prediction model was developed without any external validation, and it was rated as <u>low risk of bias for all domains</u> , consider downgrading to high risk of bias . Such a model can only be considered as low risk of bias, if the development was based on a very large data set and included some form of internal validation.
High risk of bias	If at least one domain is judged to be at high risk of bias .
Unclear risk of bias	If an unclear risk of bias was noted in at least one domain and it was low risk for all other domains.

Complete for each evaluation of a distinct model.

Reaching an overall judgment about applicability of the prediction model evaluation

Low concerns for applicability	If low concerns for applicability for all domains, the prediction model evaluation is judged to have low concerns for applicability .
High concerns for applicability	If high concerns for applicability for at least one domain, the prediction model evaluation is judged to have high concerns for applicability .
Unclear concerns for applicability	If unclear concerns (but no “high concern”) for applicability for at least one domain, the prediction model evaluation is judged to have unclear concerns for applicability overall.

Overall judgment about risk of bias and applicability of the prediction model evaluation

Overall judgment of risk of bias	RISK: (low/ high/ unclear)	Low
<i>Summary of sources of potential bias: Low</i>		
Overall judgment of applicability	CONCERN: (low/ high/ unclear)	Low
<i>Summary of applicability concerns: Low</i>		

PROBAST = Prediction model Risk Of Bias ASsessment Tool.

4. Cameli M, et al:

Step 2: Classify the type of prediction model evaluation

Classify the evaluation based on its aim			
Type of prediction study	PROBAST boxes to complete	Tick as appropriate	Definitions for type of prediction model study
Development only	Dev	X	Prediction model development without external validation. These studies may include internal validation methods such as bootstrapping and cross-validation techniques
Development and validation	Dev and Val		Prediction model development combined with external validation in other participants in the same article
Validation only	Val		External validation of existing (previously developed) model in other participants
<i>This table should be completed once for each publication being assessed and for each relevant outcome in your review.</i>			
Publication reference	Cameli M, Pastore MC, De Carli G, Henein MY, Mandoli GE, Lisi E, et al. ACUTE HF score, a multiparametric prognostic tool for acute heart failure: a real-life study, international Journal of Cardiology. https://doi.org/10.1016/j.ijcard.2019.07.015		
Models of interest	ACUTE HF		
Outcome of interest	Find the best predictors of prognosis in patients hospitalized for AHF, and to describe the predictive role of echocardiographic parameters, individually or in combination, in AHF, investigating the additive value their use could provide in daily clinical practice.		

Step 3: Assess risk of bias and applicability

DOMAIN 1: Participants			
A. Risk of Bias			
Describe the sources of data and criteria for participant selection: A retrospective study was conducted in a cohort of 771 patients hospitalized for signs and symptoms (i.e. rapid worsening of dyspnoea or peripheral oedema, cardiogenic shock, lung congestion at chest-X-ray) of AHF in the coronary care unit (CCU) of Santa Maria delle Scotte Hospital in Siena, between 2011 and 2013.			
		Dev	Val
1.1 Were appropriate data sources used, e.g. cohort, RCT or nested case-control study data?		Y	
1.2 Were all inclusions and exclusions of participants appropriate?		Y	
Risk of bias introduced by selection of participants	RISK: (low/ high/	Low	

	unclear)		
<i>Rationale of bias rating:</i>			
Inclusion and exclusion criteria were commented.			
B. Applicability			
Describe included participants, setting and dates: Of the 830 patients enrolled, 771 were eligible for the analysis. Mean age was 72.3 (± 13.5) years, 65% were male and 35% female. Only 3% had a body mass index above 30 kg/m ² . Systolic blood pressure was 125.3 $\pm$ 32.2mmHg and heart rate 86.2 $\pm$ 30.5 beats/min. 26% of patients had previous hospitalization for HF, 7% had history of stroke or transient ischemic attack (TIA), and 22% had chronic renal failure. 46% patients had AHF of ischemic etiology and 33% had a nonischemic etiology. 13% of patients received non-invasive ventilation (NIV) at the time of admission.			
Concern that the included participants and setting do not match the review question	CONCERN: (low/ high/ unclear)	Low	
<i>Rationale of applicability rating: All data were correctly specified.</i>			

DOMAIN 2: Predictors			
A. Risk of Bias			
List and describe predictors included in the final model, e.g. definition and timing of assessment: Age > 76, Creatinine > 2 mg/dl, Use of non-invasive ventilation, transient ischemic attack or stroke, Ejection fraction, Prior hospitalization for acute heart failure, Mitral dysfunction			
		De v	Val
2.1 Were predictors defined and assessed in a similar way for all participants?		Y	
2.2 Were predictor assessments made without knowledge of outcome data?		Y	
2.3 Are all predictors available at the time the model is intended to be used?		Y	
Risk of bias introduced by predictors or their assessment	RISK: (low/ high/ unclear)	Low	
<i>Rationale of bias rating: The article has a good selection of variables combining variables known as risk factors in AHF.</i>			
B. Applicability			
Concern that the definition, assessment or timing of predictors in the model do not match the review question	CONCERN: (low/ high/ unclear)	Low	
<i>Rationale of applicability rating: They use a simple group of variables, easy to get for develop one usefull and very applicable model.</i>			
DOMAIN 3: Outcome			
A. Risk of Bias			
Describe the outcome, how it was defined and determined, and the time interval between predictor assessment and outcome determination: Find the best predictors of prognosis in patients hospitalized for AHF, and to describe the predictive role of echocardiographic parameters, individually or in combination, in AHF, investigating the aditive value their use could provide in daily clinical practice.			
		Dev	Val
3.1 Was the outcome determined appropriately?		Y	
3.2 Was a pre-specified or standard outcome definition used?		Y	

3.3 Were predictors excluded from the outcome definition?	Y	
3.4 Was the outcome defined and determined in a similar way for all participants?	Y	
3.5 Was the outcome determined without knowledge of predictor information?	Y	
3.6 Was the time interval between predictor assessment and outcome determination appropriate?	Y	
Risk of bias introduced by the outcome or its determination	RISK: (low/ high/ unclear)	Low
<i>Rationale of bias rating: It was very well defined outcome, with good model</i>		
B. Applicability		
<i>At what time point was the outcome determined: at one time in the future</i>		
<i>If a composite outcome was used, describe the relative frequency/distribution of each contributing outcome:</i>		
Concern that the outcome, its definition, timing or determination do not match the review question	CONCERN: (low/ high/ unclear)	Low
<i>Rationale of applicability rating: It was correctly presented the information.</i>		
DOMAIN 4: Analysis		
Risk of Bias		
<p><i>Describe numbers of participants, number of candidate predictors (for DEV only), outcome events and events per candidate predictor (for DEV only): 771 were eligible for the analysis from 830. age; ischemic etiology, non-ischemic etiology, previous ICD implantation, and the use of dopamine. Among the echocardiographic variables, IVS, PW, LV EDD, ESD, left atrial area and LV EF.</i></p>		
<p><i>Describe how the model was developed (predictor selection, optimism, risk groups, model performance): Analyses were performed using the SPSS (Statistical Package for the Social Sciences, Chicago, Illinois) software release 12.0. Binary variables are presented as counts and percentages, while numerical variables are presented as mean $\pm$ SD or, in the case of high skewness and/or non-normality, as median [1st quartile, 3rd quartile]. Basic univariate comparisons were performed using standard Chi-square test in the case of categorical comparisons, Student t-test in case of normal comparisons and Mann-Whitney U non-parametric test for non-normal comparisons. Normality of numerical variables was determined by the Kolmogorov-Smirnov test. For each variable, the relationship with the length of stay in CCU and in ordinary regimen was evaluated using univariate and multivariate analyses. Multivariate fit for hospitalization days was performed via a generalized linear model (GLM) with Poisson response. Variables with >100 missing values were excluded from the model. When the Pearson correlation between variables was >0.50, only the variable with the best fit was left in the model. Significance level was set at 0.05 before performing a standard Bonferroni correction. The predictive power of the variables for mortality at 30 days, 6 months and 5 years was analyzed using the univariate and multivariate analysis. The multivariate analysis, as for the duration of hospitalization, was applied in a first phase only to the variables that presented missing data in <100 patients. In a second phase, the analysis was also extended to some variables previously excluded (glucose, urea, EDD, ESD, LA area, TAPSE, right EDD and sPAP). The most significant variables according to the 3 mortality multivariate models (beta values between 0,5 and 1,5; $p < 0.005$) and the variables with high univariate prediction power ($p < 0.00001$) and few missing values (<100) were the selected variables to produce the predictive score. Among this group of selected variables, the numerical predictors were discretized using a univariate Youden index for balancing sensitivity and specificity (applied on the binary 6-months mortality). The discretized variables were then used to compute the score using a simple binomial GLM. Linear estimators from this model were used as weights for each of the discretized variables, to create the new score as the weighted sum of the discretized variables. Receiver operating characteristics (ROC) and area under curve (AUC) were composed to assess the score's predictive power for 30- day, 6-month and 5-year mortality rates and to find optimal cut-off values</i></p>		

for each variable. The score was used to classify patients into 3 groups and Kaplan-Meier survival curves were computed for these groups and compared with the survival curves of the score's numerical components. All analyses included age, sex, BMI, BP, HR and smoker indicators as confounding variables.			
Describe whether and how the model was validated, either internally (e.g. bootstrapping, cross validation, random split sample) or externally (e.g. temporal validation, geographical validation, different setting, different type of participants): They develop the score by internal validation.			
Describe the performance measures of the model, e.g. (re)calibration, discrimination, (re)classification, net benefit: Receiver Operating Characteristic (ROC) curves with area under the curves (AUC) were calculated to assess the predictive power of the new score for 30-day, 6-months and 5-years mortality. The AUC value was 0,78; 0,79; and 0,76 respectively. The ACUTE HF score proved to have better prognostic value than each variable included in the score (age, creatinine, NIV, stroke/TIA, LV EF, hospitalization, MR) individually. Moreover, ACUTE HF score proved to be superior than the ADHERE score, previously applied in AHF, based on the ROC produced sensitivity and specificity for mortality prediction. Kaplan-Meier survival curves showed that the ACUTE HF score risk stratification was superior than single variables, e.g. LV EF and creatinine, showing clear differences in free events survival rates between the three risk groups			
Describe any participants who were excluded from the analysis: active neoplasms; previous heart transplantation or left ventricular assist device implantation.			
Describe missing data on predictors and outcomes as well as methods used for missing data: Variables with N100 missing values were excluded from the model.			
		Dev	Val
4.1 Were there a reasonable number of participants with the outcome?		Y	
4.2 Were continuous and categorical predictors handled appropriately?		Y	
4.3 Were all enrolled participants included in the analysis?		Y	
4.4 Were participants with missing data handled appropriately?		Y	
4.5 Was selection of predictors based on univariable analysis avoided?		Y	
4.6 Were complexities in the data (e.g. censoring, competing risks, sampling of controls) accounted for appropriately?		NI	
4.7 Were relevant model performance measures evaluated appropriately?		Y	
4.8 Were model overfitting and optimism in model performance accounted for?		Y	
4.9 Do predictors and their assigned weights in the final model correspond to the results from the reported multivariable analysis?		Y	
Risk of bias introduced by the analysis	RISK: (low/unclear) high/	Low	
Rationale of bias rating: Describe how they get the score			

Step 4: Overall assessment

Use the following tables to reach overall judgments about risk of bias and concerns for applicability of the prediction model evaluation (development and/or validation) across all assessed domains.

Reaching an overall judgment about risk of bias of the prediction model evaluation	
Low risk of bias	If all domains were rated low risk of bias. If a prediction model was developed without any external validation, and it was rated as <u>low risk of bias for all domains</u> , consider downgrading to high risk of bias . Such a model can only be considered as low risk of bias, if the development was based on a very large data set and included some form of internal validation.
High risk of bias	If at least one domain is judged to be at high risk of bias .
Unclear risk of bias	If an unclear risk of bias was noted in at least one domain and it was low risk for all other domains.

Complete for each evaluation of a distinct model.

Reaching an overall judgment about applicability of the prediction model evaluation	
Low concerns for applicability	If low concerns for applicability for all domains, the prediction model evaluation is judged to have low concerns for applicability .
High concerns for applicability	If high concerns for applicability for at least one domain, the prediction model evaluation is judged to have high concerns for applicability .
Unclear concerns for applicability	If unclear concerns (but no “high concern”) for applicability for at least one domain, the prediction model evaluation is judged to have unclear concerns for applicability overall.

Overall judgment about risk of bias and applicability of the prediction model evaluation		
Overall judgment of risk of bias	RISK: (low/ high/ unclear)	Low
<i>Summary of sources of potential bias: Low</i>		
Overall judgment of applicability	CONCERN: (low/ high/ unclear)	Low
<i>Summary of applicability concerns: Low</i>		

PROBAST = Prediction model Risk Of Bias ASsessment Too

5. Chen YJ, et al (1):

Step 2: Classify the type of prediction model evaluation

Classify the evaluation based on its aim			
Type of prediction study	PROBAST boxes to complete	Tick as appropriate	Definitions for type of prediction model study
Development only	Dev		Prediction model development without external validation. These studies may include internal validation methods such as bootstrapping and cross-validation techniques
Development and validation	Dev and Val		Prediction model development combined with external validation in other participants in the same article
Validation only	Val	X	External validation of existing (previously developed) model in other participants
<i>This table should be completed once for each publication being assessed and for each relevant outcome in your review.</i>			
Publication reference	Chen YJ, Sung SH, Cheng HM, Huang WM, Wu CL, Huang CJ, et al. Performance of AHEAD Score in an Asian Cohort of Acute Heart Failure With Either Preserved or Reduced Left Ventricular Systolic Function. J Am Heart Assoc. 2017;6:e004297. DOI: 10.1161/JAHA.116.004297.		
Models of interest	Performance of AHEAD Score in an Asian Cohort of Acute Heart Failure With Either Preserved or Reduced Left Ventricular Systolic Function		
Outcome of interest	Performance of AHEAD score in an Asian cohort of acute heart failure with patients with either preserved or reduced left ventricular systolic function.		

Step 3: Assess risk of bias and applicability

DOMAIN 1: Participants			
A. Risk of Bias			
<i>Describe the sources of data and criteria for participant selection: The study population was drawn from the heart failure registry of Taipei Veterans General Hospital (HARVEST registry) that included patients hospitalized for AHF, defined as new-onset or gradually or rapidly worsening heart failure symptoms and signs requiring urgent therapy. Consecutive AHF patients with New York Heart Association functional class III or IV symptoms, compatible presentations of chest radiograph, and responses to diuretics, were enrolled. (only enrolled subjects who had undergone echocardiographic examinations in the analysis, selection bias was not avoidable.)</i>			
		Dev	Val
1.1 Were appropriate data sources used, e.g. cohort, RCT or nested case-control study data?			Y
1.2 Were all inclusions and exclusions of participants appropriate?			N
Risk of bias introduced by selection of participants	RISK: (low/ high/ unclear)		high
<i>Rationale of bias rating: By only including patients with echocardiography, they committed selection bias.</i>			
B. Applicability			
<i>Describe included participants, setting and dates: In a total of 2663 eligible subjects hospitalized primarily for AHF from October 2003 to December 2012, 175 patients with complete data of hematology, biochemistry, and echocardiogram constituted this study population. A total of 175 patients (age 69.9±15.4years, 77.1% men, 67.4% HFrEF) constituted this study, and they were characterized by multiple comorbidities: 74.3% had hypertension, 50.3% had diabetes mellitus, 58.3% had coronary artery disease, and 26.9% had atrial fibrillation.</i>			

Concern that the included participants and setting do not match the review question	CONCERN: (low/ high/ unclear)		Low
<i>Rationale of applicability rating: The classification of the participants in AHF was under the correct criteria, and they agree with the definition of the pathology in the research question.</i>			

DOMAIN 2: Predictors			
A. Risk of Bias			
<p>List and describe predictors included in the final model, e.g. definition and timing of assessment: 1 point for each of A: atrial fibrillation, H: hemoglobin <130 g/L for men and 120 g/L for women, E: elderly (age >70 years), A: abnormal renal parameters (creatinine >130 μmol/L), and D: diabetes mellitus. By a cut-off value of 8.6 mg/dL derived from receiver operating characteristic curve analysis, uric acid was incorporated with the AHEAD score to construct the AHEAD-U index. Hemogram, renal function, serum electrolytes, and N-terminal pro-brain natriuretic peptide (NT-proBNP) were measured immediately upon presentation to the hospital. Uric acid and the lipid profiles were obtained at fasting the next morning. Estimated glomerular filtration rate was then calculated using the modified glomerular filtration rate estimating equation for Chinese patients. Comorbidities were identified by medical history and the associated measures during the index hospitalization. Therefore, atrial fibrillation was diagnosed based on past or ad-hoc ECGs. Diabetes mellitus was confirmed by medical records and a glycosylated hemoglobin ≥6.5% during the index hospitalization. Hypertension was defined according to medical records.</p>			
		De v	Val
2.1 Were predictors defined and assessed in a similar way for all participants?			Y
2.2 Were predictor assessments made without knowledge of outcome data?			Y
2.3 Are all predictors available at the time the model is intended to be used?			Y
Risk of bias introduced by predictors or their assessment	RISK: (low/ high/ unclear)		Low
<i>Rationale of bias rating: They describe the variables included in each predictive model, with a clear definition, which were evaluated during the follow-up, at the appropriate time.</i>			
B. Applicability			
Concern that the definition, assessment or timing of predictors in the model do not match the review question	CONCERN: (low/ high/ unclear)		Low
<i>Rationale of applicability rating: Biomarkers can be applied in a correct time suitable for risk classification.</i>			

DOMAIN 3: Outcome
A. Risk of Bias

<p><i>Describe the outcome, how it was defined and determined, and the time interval between predictor assessment and outcome determination:</i></p> <p>Describe the outcome, how it was defined and determined, and the time interval between predictor assessment and outcome determination: Performance of AHEAD score in an Asian cohort of acute heart failure with patients with either preserved or reduced left ventricular systolic function. The study may extend the clinical applications of risk-predicting models(Predicting 3-Year all-cause mortality and cardiovascular mortality) in patients with HFpEF, while the established prognostic calculators were constructed based on study populations, in which the majority were HFrEF. The clinical end point was the occurrence of all-cause mortality or cardiovascular death. Mortality in the study population was ascertained by the International Classification of Diseases, 9th Revision coding in the National Death Registry with a follow-up duration of up to 3 years..</p>			
		Dev	Val
3.1 Was the outcome determined appropriately?			Y
3.2 Was a pre-specified or standard outcome definition used?			Y
3.3 Were predictors excluded from the outcome definition?			Y
3.4 Was the outcome defined and determined in a similar way for all participants?			Y
3.5 Was the outcome determined without knowledge of predictor information?			Y
3.6 Was the time interval between predictor assessment and outcome determination appropriate?			Y
Risk of bias introduced by the outcome or its determination		RISK: (low/ high/ unclear)	Low
<p><i>Rationale of bias rating: Clear definition of the desired outcome, in addition to correct monitoring.</i></p>			
B. Applicability			
<p><i>At what time point was the outcome determined:</i></p> <p>It was determined during the follow-up in the future.</p>			
<p><i>If a composite outcome was used, describe the relative frequency/distribution of each contributing outcome:</i></p>			
Concern that the outcome, its definition, timing or determination do not match the review question		CONCERN: (low/ high/ unclear)	Low
<p><i>Rationale of applicability rating: The time in which the outcome is determined is consistent with the proposed research question.</i></p>			

DOMAIN 4: Analysis			
Risk of Bias			
Describe numbers of participants, number of candidate predictors (for DEV only), outcome events and events per candidate predictor (for DEV only): The population included in the study is 175. Candidate predictors are not reported, so the EPV cannot be estimated			
Describe how the model was developed (predictor selection, optimism, risk groups, model performance): First, the AHEAD model was used in the Asian population, after that, uric acid was added as a predictor value to the ADHERE study, and the same population was used for internal validation. Compared with the AHEAD score, AHEAD-U significantly reclassified more patients into the correct risk categories and achieved a net reclassification improvement of 19.66% (P=0.0002) in all-cause mortality and 20.08% (P=0.0025) in cardiovascular mortality. the risk groups were divided with respect to the score taken used in the AHEAD model			
Describe whether and how the model was validated, either internally (e.g. bootstrapping, cross validation, random split sample) or externally (e.g. temporal validation, geographical validation, different setting, different type of participants): not mentioned			
Describe the performance measures of the model, e.g. (re)calibration, discrimination, (re)classification, net benefit: Discrimination: Compared with the AHEAD score, AHEAD-U significantly reclassified more patients into the correct risk categories and achieved a net reclassification improvement of 19.66% (P=0.0002) in all-cause mortality and 20.08% (P=0.0025) in cardiovascular mortality. the risk groups were divided with respect to the score taken used in the AHEAD model. there are no data on discrimination or calibration			
Describe any participants who were excluded from the analysis: Subjects with acute coronary syndrome, severe hepatic disease, or severe infection were excluded from this analysis.			
Describe missing data on predictors and outcomes as well as methods used for missing data: not mentioned			
		Dev	Val
4.1 Were there a reasonable number of participants with the outcome?			PY
4.2 Were continuous and categorical predictors handled appropriately?			Y
4.3 Were all enrolled participants included in the analysis?			Y
4.4 Were participants with missing data handled appropriately?			PY
4.5 Was selection of predictors based on univariable analysis avoided?			NI
4.6 Were complexities in the data (e.g. censoring, competing risks, sampling of controls) accounted for appropriately?			NI
4.7 Were relevant model performance measures evaluated appropriately?			PN
4.8 Were model overfitting and optimism in model performance accounted for?			NI
4.9 Do predictors and their assigned weights in the final model correspond to the results from the reported multivariable analysis?			Y
Risk of bias introduced by the analysis	RISK: (low/ high/ unclear)		High
Rationale of bias rating: does not comment on much important information, such as calibration or discrimination, in addition to not presenting data on the number of events in the cohort study			

Step 4: Overall assessment

Use the following tables to reach overall judgments about risk of bias and concerns for applicability of the prediction model evaluation (development and/or validation) across all assessed domains.

Reaching an overall judgment about risk of bias of the prediction model evaluation	
Low risk of bias	If all domains were rated low risk of bias. If a prediction model was developed without any external validation, and it was rated as low risk of bias for all domains , consider downgrading to high risk of bias . Such a model can only be considered as low risk of bias, if the development was based on a very large data set and included some form of internal validation.
High risk of bias	If at least one domain is judged to be at high risk of bias .
Unclear risk of bias	If an unclear risk of bias was noted in at least one domain and it was low risk for all other domains.

Complete for each evaluation of a distinct model.

Reaching an overall judgment about applicability of the prediction model evaluation	
Low concerns for applicability	If low concerns for applicability for all domains, the prediction model evaluation is judged to have low concerns for applicability .
High concerns for applicability	If high concerns for applicability for at least one domain, the prediction model evaluation is judged to have high concerns for applicability .
Unclear concerns for applicability	If unclear concerns (but no “high concern”) for applicability for at least one domain, the prediction model evaluation is judged to have unclear concerns for applicability overall.

Overall judgment about risk of bias and applicability of the prediction model evaluation		
Overall judgment of risk of bias	RISK: (low/ high/ unclear)	
Summary of sources of potential bias: High		
Overall judgment of applicability	CONCERN: (low/ high/ unclear)	
Summary of applicability concerns: Low		

PROBAST = Prediction model Risk Of Bias ASsessment Tool.

6. Chen YJ, et al (2):

Step 2: Classify the type of prediction model evaluation

Classify the evaluation based on its aim			
Type of prediction study	PROBAST boxes to complete	Tick as appropriate	Definitions for type of prediction model study
Development only	Dev	X	Prediction model development without external validation. These studies may include internal validation methods such as bootstrapping and cross-validation techniques
Development and validation	Dev and Val		Prediction model development combined with external validation in other participants in the same article
Validation only	Val		External validation of existing (previously developed) model in other participants
<i>This table should be completed once for each publication being assessed and for each relevant outcome in your review.</i>			
Publication reference	Chen YJ, Sung SH, Cheng HM, Huang WM, Wu CL, Huang CJ, et al. Performance of AHEAD Score in an Asian Cohort of Acute Heart Failure With Either Preserved or Reduced Left Ventricular Systolic Function. J Am Heart Assoc. 2017;6:e004297. DOI: 10.1161/JAHA.116.004297.		
Models of interest	Performance of AHEAD Score in an Asian Cohort of Acute Heart Failure With Either Preserved or Reduced Left Ventricular Systolic Function		
Outcome of interest	Predicting 3-Year all-cause mortality and cardiovascular mortality		

Step 3: Assess risk of bias and applicability

DOMAIN 1: Participants			
A. Risk of Bias			
<i>Describe the sources of data and criteria for participant selection: The study population was drawn from the heart failure registry of Taipei Veterans General Hospital (HARVEST registry) that included patients hospitalized for AHF, defined as new-onset or gradually or rapidly worsening heart failure symptoms and signs requiring urgent therapy. Consecutive AHF patients with New York Heart Association functional class III or IV symptoms, compatible presentations of chest radiograph, and responses to diuretics, were enrolled. (only enrolled subjects who had undergone echocardiographic examinations in the analysis, selection bias was not avoidable.)</i>			
		De v	Val
1.1 Were appropriate data sources used, e.g. cohort, RCT or nested case-control study data?		Y	
1.2 Were all inclusions and exclusions of participants appropriate?		N	
Risk of bias introduced by selection of participants	RISK: (low/ high/ unclear)	high	
<i>Rationale of bias rating: By only including patients with echocardiography, they committed selection bias.</i>			
B. Applicability			
<i>Describe included participants, setting and dates: In a total of 2663 eligible subjects hospitalized primarily for AHF from October 2003 to December 2012, 175 patients with complete data of hematology, biochemistry, and echocardiogram constituted this study population. A total of 175 patients (age 69.9±15.4years, 77.1% men, 67.4% HFrEF) constituted this study, and they were characterized by multiple comorbidities: 74.3% had</i>			

<i>hypertension, 50.3% had diabetes mellitus, 58.3% had coronary artery disease, and 26.9% had atrial fibrillation.</i>			
Concern that the included participants and setting do not match the review question	CONCERN: (low/ high/ unclear)	Low	
<i>Rationale of applicability rating: The classification of the participants in AHF was under the correct criteria, and they agree with the definition of the pathology in the research question.</i>			

DOMAIN 2: Predictors			
A. Risk of Bias			
<i>List and describe predictors included in the final model, e.g. definition and timing of assessment: 1 point for each of A: atrial fibrillation, H: hemoglobin <130 g/L for men and 120 g/L for women, E: elderly (age >70 years), A: abnormal renal parameters (creatinine >130 μmol/L), and D: diabetes mellitus. By a cut-off value of 8.6 mg/dL derived from receiver operating characteristic curve analysis, uric acid was incorporated with the AHEAD score to construct the AHEAD-U index. Hemogram, renal function, serum electrolytes, and N-terminal pro-brain natriuretic peptide (NT-proBNP) were measured immediately upon presentation to the hospital. Uric acid and the lipid profiles were obtained at fasting the next morning. Estimated glomerular filtration rate was then calculated using the modified glomerular filtration rate estimating equation for Chinese patients. Comorbidities were identified by medical history and the associated measures during the index hospitalization. Therefore, atrial fibrillation was diagnosed based on past or ad-hoc ECGs. Diabetes mellitus was confirmed by medical records and a glycosylated hemoglobin ≥6.5% during the index hospitalization. Hypertension was defined according to medical records.</i>			
		De v	Val
2.1 Were predictors defined and assessed in a similar way for all participants?		Y	
2.2 Were predictor assessments made without knowledge of outcome data?		Y	
2.3 Are all predictors available at the time the model is intended to be used?		Y	
Risk of bias introduced by predictors or their assessment	RISK: (low/ high/ unclear)	Low	
<i>Rationale of bias rating: They describe the variables included in each predictive model, with a clear definition, which were evaluated during the follow-up, at the appropriate time.</i>			
B. Applicability			
Concern that the definition, assessment or timing of predictors in the model do not match the review question	CONCERN: (low/ high/ unclear)	Low	
<i>Rationale of applicability rating: Biomarkers can be applied in a correct time suitable for risk classification.</i>			

DOMAIN 3: Outcome
A. Risk of Bias

<p><i>Describe the outcome, how it was defined and determined, and the time interval between predictor assessment and outcome determination:</i></p> <p>Predict the risk of any cause of death or cardiac death in a 3-year follow-up. Adding uric acid as a predictor variable to the AHEAD model. The authors report that the importance of uric acid in predicting mortality in previous studies had been demonstrated, so they decided to create a new model.</p>			
		Dev	Val
3.1 Was the outcome determined appropriately?		Y	
3.2 Was a pre-specified or standard outcome definition used?		Y	
3.3 Were predictors excluded from the outcome definition?		Y	
3.4 Was the outcome defined and determined in a similar way for all participants?		Y	
3.5 Was the outcome determined without knowledge of predictor information?		Y	
3.6 Was the time interval between predictor assessment and outcome determination appropriate?		Y	
Risk of bias introduced by the outcome or its determination		RISK: (low/ high/ unclear)	Low
<p><i>Rationale of bias rating: Clear definition of the desired outcome, in addition to correct monitoring.</i></p>			
B. Applicability			
<p><i>At what time point was the outcome determined:</i></p> <p><i>It was determined during the follow-up in the future.</i></p> <p><i>If a composite outcome was used, describe the relative frequency/distribution of each contributing outcome:</i></p>			
Concern that the outcome, its definition, timing or determination do not match the review question		CONCERN: (low/ high/ unclear)	Low
DOMAIN 4: Analysis			
Risk of Bias			
<p><i>Describe numbers of participants, number of candidate predictors (for DEV only), outcome events and events per candidate predictor (for DEV only): The population included in the study is 175. Candidate predictors are not reported, so the EPV cannot be estimated</i></p>			
<p><i>Describe how the model was developed (predictor selection, optimism, risk groups, model performance): First, the AHEAD model was used in the Asian population, after that, uric acid was added as a predictor value to the ADHERE study, and the same population was used for internal validation. Compared with the AHEAD score, AHEAD-U significantly reclassified more patients into the correct risk categories and achieved a net reclassification improvement of 19.66% (P=0.0002) in all-cause mortality and 20.08% (P=0.0025) in cardiovascular mortality. the risk groups were divided with respect to the score taken used in the AHEAD model</i></p>			
<p><i>Describe whether and how the model was validated, either internally (e.g. bootstrapping, cross validation, random split sample) or externally (e.g. temporal validation, geographical validation, different setting, different type of participants): not mentioned</i></p>			
<p><i>Describe the performance measures of the model, e.g. (re)calibration, discrimination, (re)classification, net benefit: Discrimination: Compared with the AHEAD score, AHEAD-U significantly reclassified more patients into the correct risk categories and achieved a net reclassification improvement of 19.66% (P=0.0002) in all-cause mortality and 20.08% (P=0.0025) in cardiovascular mortality. the risk groups were divided with respect to the score taken used in the AHEAD model. there are no data on discrimination or calibration</i></p>			
<p><i>Describe any participants who were excluded from the analysis: Subjects with acute coronary syndrome, severe hepatic disease, or severe infection were excluded from this analysis.</i></p>			
<p><i>Describe missing data on predictors and outcomes as well as methods used for missing data: not mentioned</i></p>			

		Dev	Val
4.1	Were there a reasonable number of participants with the outcome?	PY	
4.2	Were continuous and categorical predictors handled appropriately?	Y	
4.3	Were all enrolled participants included in the analysis?	Y	
4.4	Were participants with missing data handled appropriately?	PY	
4.5	Was selection of predictors based on univariable analysis avoided?	NI	
4.6	Were complexities in the data (e.g. censoring, competing risks, sampling of controls) accounted for appropriately?	NI	
4.7	Were relevant model performance measures evaluated appropriately?	PN	
4.8	Were model overfitting and optimism in model performance accounted for?	NI	
4.9	Do predictors and their assigned weights in the final model correspond to the results from the reported multivariable analysis?	Y	
Risk of bias introduced by the analysis		RISK: (low/ high/ unclear)	High
<i>Rationale of bias rating:</i> <i>does not comment on much important information, such as calibration or discrimination, in addition to not presenting data on the number of events in the cohort study</i>			
<i>Rationale of applicability rating: The time in which the outcome is determined is consistent with the proposed research question.</i>			

Step 4: Overall assessment

Use the following tables to reach overall judgments about risk of bias and concerns for applicability of the prediction model evaluation (development and/or validation) across all assessed domains.	
Reaching an overall judgment about risk of bias of the prediction model evaluation	
Low risk of bias	If all domains were rated low risk of bias. If a <u>prediction model was developed without any external validation</u> , and it was rated as <u>low risk of bias for all domains</u> , consider downgrading to high risk of bias . Such a model can only be considered as low risk of bias, if the development was based on a very large data set and included some form of internal validation.
High risk of bias	If at least one domain is judged to be at high risk of bias .
Unclear risk of bias	If an unclear risk of bias was noted in at least one domain and it was low risk for all other domains.
<i>Complete for each evaluation of a distinct model.</i>	
Reaching an overall judgment about applicability of the prediction model evaluation	
Low concerns for applicability	If low concerns for applicability for all domains, the prediction model evaluation is judged to have low concerns for applicability .
High concerns for applicability	If high concerns for applicability for at least one domain, the prediction model evaluation is judged to have high concerns for applicability .
Unclear concerns for applicability	If unclear concerns (but no “high concern”) for applicability for at least one domain, the prediction model evaluation is judged to have unclear concerns for applicability .

applicability	overall.
Overall judgment about risk of bias and applicability of the prediction model evaluation	
Overall judgment of risk of bias	RISK: (low/ high/ unclear)
Summary of sources of potential bias: High	
Overall judgment of applicability	CONCERN: (low/ high/ unclear)
Summary of applicability concerns: Low	

PROBAST = Prediction model Risk Of Bias ASsessment Tool.

7. Collins, SP, et al:

Step 2: Classify the type of prediction model evaluation

Classify the evaluation based on its aim			
Type of prediction study	PROBAST boxes to complete	Tick as appropriate	Definitions for type of prediction model study
Development only	Dev	X	Prediction model development without external validation. These studies may include internal validation methods such as bootstrapping and cross-validation techniques
Development and validation	Dev and Val		Prediction model development combined with external validation in other participants in the same article
Validation only	Val		External validation of existing (previously developed) model in other participants
<i>This table should be completed once for each publication being assessed and for each relevant outcome in your review.</i>			
Publication reference	Collins SP, Jenkins CA, Harrell FE, Liu D, Miller KF, Lindsell CJ, et al. Identification of Emergency Department Patients With Acute Heart Failure at Low Risk for 30-Day Adverse Events: The STRATIFY. JACC Heart Fail. 2015 October ; 3(10): 737–747. doi:10.1016/j.jchf.2015.05.007.		
Models of interest	Identification of Emergency Department Patients With Acute Heart Failure at Low Risk for 30-Day Adverse Events: The STRATIFY Decision Tool		
Outcome of interest	No prospectively derived or validated decision tools identify emergency department (ED) patients with acute heart failure (AHF) at low risk for 30-day adverse events who are thus potential candidates for safe ED discharge		

Step 3: Assess risk of bias and applicability

DOMAIN 1: Participants			
A. Risk of Bias			
<p><i>Describe the sources of data and criteria for participant selection: We conducted a prospective, observational cohort study, STRATIFY (Improving Heart Failure Risk Stratification in the ED), from July 20, 2007, to February 4, 2011, at 2 university-affiliated tertiary care EDs and 2 community EDs. Our study recruited patients at 1 ED in Nashville, Tennessee, and 3 EDs in Cincinnati, Ohio. ED patients were screened for possible AHF by use of modified Framingham criteria (25). Our modification reflects contemporary practice for making a preliminary diagnosis of AHF. Three Framingham criteria were not used: 1) circulation time; 2) vital capacity; and 3) weight loss in response to treatment. Vital capacity and circulation time are not typically available in the acute setting, and weight loss in response to treatment would only assist with a retrospective diagnosis. Inclusion of patients required the presence of at least 2 major, or 1 major and 2 minor, modified Framingham criteria (major: paroxysmal nocturnal dyspnea, neck vein distention, rales, cardiomegaly on chest radiograph, acute pulmonary edema, S3 gallop, or hepatjugular reflux; minor: ankle edema, night cough, dyspnea on exertion, hepatomegaly, pleural effusion, or tachycardia ≥ 120 beats/min). Patients had to be willing and able to give informed consent and be at least 18 years of age. Although use of B-type natriuretic peptide (BNP) <100 pg/ml to support a non-AHF diagnosis is standard at our institutions, it was not used for exclusion in the proposed study, because lack of BNP elevation was expected to exclude some low-risk patients. Although use of BNP to enroll patients may have increased the probability of AHF, the investigators did not want to exclude subjects who were clinically believed to have AHF because of a false negative BNP result.</i></p>			
		De v	Val
1.1 Were appropriate data sources used, e.g. cohort, RCT or nested case-control study data?		Y	
1.2 Were all inclusions and exclusions of participants appropriate?		Y	
Risk of bias introduced by selection of participants	RISK: (low/ high/ unclear)	low	
<i>Rationale of bias rating: Applied correct criteria for inclusion and exclusion of participants</i>			
B. Applicability			
<p><i>Describe included participants, setting and dates: A total of 2,074 subjects with signs and symptoms of AHF were recruited. Sixty-three withdrew, and 18 were lost to follow-up. Of the remaining cohort, 1,033 were identified by the study investigator's review as having AHF in the ED. There were 94 charts reviewed in duplicate by the ED reviewers, which resulted in a kappa of 0.90 (95% confidence interval [CI]: 0.80 to 1.00). Of the 1,033 patients included in our cohort, our cardiology oversight group adjudicated 820 patient charts and determined 762 (93%) had a diagnosis of AHF. With the exception of ejection fraction (9.2%), none of the variables had more than 4% missingness. The median age was 64 years, 57% were male, and 44% were African American. Overall, patient comorbidities were consistent with a chronic HF population, with 74% having a prior history of HF, 35% with prior myocardial infarction, and 22% with a history of renal disease. As a group, renal function was impaired (median estimated glomerular filtration rate 52.8 ml/min/1.73 m²), systolic blood pressure was elevated (median 144 mm Hg), and BNP levels were elevated (median 994 pg/ml).</i></p>			
Concern that the included participants and setting do not match the review question	CONCERN: (low/ high/ unclear)	Low	
<i>Rationale of applicability rating: The classification of the participants in AHF was under the correct criteria, and they agree with the definition of the pathology in the research question.</i>			
DOMAIN 2: Predictors			
A. Risk of Bias			
<p><i>List and describe predictors included in the final model, e.g. definition and timing of assessment: Age, body mass index, B-type natriuretic peptide, blood urea nitrogen, diastolic blood pressure, Sodium, respiratory rate, arterial oxygen saturation, Troponin I (cubic root), Dialysis, On supplemental O₂, On outpatient angiotensin-converting enzyme inhibitor, QRS duration. Research assistants collected data by direct questioning of the</i></p>			

<p>patient and treating physician, as well as by a review of the electronic medical record during the first 3 h of ED management. The principal investigator or study coordinator reviewed and confirmed the accuracy of the data recorded by the research assistants. The assessors were blinded to the inpatient medical record. If the assessors agreed the ED visit was definitely or definitively not AHF-related, no further assessment was performed. For all others, a third assessor (G.J.F.) adjudicated. A subset of medical records were reviewed in duplicate to determine abstractor agreement.</p>			
		Dev	Val
2.1 Were predictors defined and assessed in a similar way for all participants?		Y	
2.2 Were predictor assessments made without knowledge of outcome data?		Y	
2.3 Are all predictors available at the time the model is intended to be used?		Y	
Risk of bias introduced by predictors or their assessment	RISK: (low/ high/ unclear)	Low	
<p><i>Rationale of bias rating: It has a complete description of the list of selected predictor variables, as well as the way in which they were studied in patients.</i></p>			
B. Applicability			
Concern that the definition, assessment or timing of predictors in the model do not match the review question	CONCERN: (low/ high/ unclear)	Low-	
<p><i>Rationale of applicability rating: The variables included can be studied in a timely manner within a predictive model of AHF.</i></p>			

DOMAIN 3: Outcome			
A. Risk of Bias			
<p><i>Describe the outcome, how it was defined and determined, and the time interval between predictor assessment and outcome determination:</i> The primary outcome was ordinal and represented the most severe adverse event experienced within 30 days of ED evaluation (Table 1). The ordinal scale was determined a priori on the basis of severity and created by the investigators, comprising both emergency physicians and cardiologists. The scale focused on events thought to be of most interest to physicians when considering ED discharge, including risk of death, ACS, cardiopulmonary resuscitation, mechanical cardiac support, mechanical ventilation, emergent dialysis, and emergency revascularization. An ordinal rather than binary outcome was used because the proportional odds model allows for parsimonious modeling of an ordinal outcome with increased power and precision compared with a binary logistic model. This hierarchy does not weight a more serious event over a less serious event in the regression model; it merely assigns the most serious outcome as the outcome of interest. If a subject experiences both an ACS and death, death is assigned as the outcome of interest. Adverse events were collected in a 2-step process. First, a chart review was performed to ascertain events that occurred during the index admission or return visits in the 30 days after enrollment. Second, phone follow-up of all patients was performed at 5 days and 30 days after enrollment.</p>			
		Dev	Val
3.1 Was the outcome determined appropriately?		Y	
3.2 Was a pre-specified or standard outcome definition used?		Y	
3.3 Were predictors excluded from the outcome definition?		Y	
3.4 Was the outcome defined and determined in a similar way for all participants?		Y	
3.5 Was the outcome determined without knowledge of predictor information?		Y	
3.6 Was the time interval between predictor assessment and outcome determination appropriate?		Y	
Risk of bias introduced by the outcome or its determination	RISK: (low/ high/ unclear)	Low	

<i>Rationale of bias rating: It has an adequate outcome definition that matches the research question in our review.</i>			
B. Applicability			
<p><i>At what time point was the outcome determined: It was determined during the follow-up in the future.</i></p> <p><i>If a composite outcome was used, describe the relative frequency/distribution of each contributing outcome:</i></p>			
Concern that the outcome, its definition, timing or determination do not match the review question	CONCERN: (low/ high/ unclear)	Low	
<i>Rationale of applicability rating: The time in which the outcome is determined is consistent with the proposed research question.</i>			

DOMAIN 4: Analysis
Risk of Bias
<p><i>Describe numbers of participants, number of candidate predictors (for DEV only), outcome events and events per candidate predictor (for DEV only): The population included in the study is 1033. Events: 126 Candidate predictors are 57, so the EPV is 2.21.</i></p>
<p><i>Describe how the model was developed (predictor selection, optimism, risk groups, model performance):</i> The STRATIFY decision tool was developed according to established strategies. First, we selected a large pool of candidate predictors based on clinical relevance and availability within the index ED visit. Then we evaluated descriptive statistics of candidate predictors for degree of missingness and level of information provided. A priori, predictors with missingness >90% were removed from consideration because their inclusion would minimize the usefulness of the tool. Categorical predictors with >95% prevalence for 1 level were also removed because of lack of variation and sufficient information to the tool. Descriptive statistics on the remaining potential predictors were calculated with median (interquartile range) or percentage (n), as appropriate. Missing data on remaining candidate predictors were imputed with single imputation that allowed for nonlinear transformations on the data to make optimal use of partial information recorded for each subject. The challenge in the development of the STRATIFY decision tool lay in the presence of a large pool of 57 candidate predictor variables that produced “noise” during traditional stepwise model selection methods. To mitigate the effects of noisy features in such a high-dimensional setting, a modeling approximation method called pre-conditioning was used for model selection. Pre-conditioning was developed to handle high dimensional data problems, in which the number of predictors was large relative to the number of events. The method has 2 steps: 1) a continuous “pre-conditioned” outcome that characterized the underlying distribution of the 6-level ordinal outcome was derived as a linear combination of all 57 candidate predictors by fitting the proportional odds model on the ordinal outcome; and 2) the best set of predictors was selected from the pool on the basis of maximizing the Akaike information criterion with backward model selection using the pre-conditioned outcome . The STRATIFY decision tool was derived by fitting a proportional odds model to the 6-level ordinal outcome using 13 variables selected in the second step of the preconditioning method. Regression splines were used in the risk prediction model. Relaxing the linearity assumption with splines allows the coefficient to vary based on the covariate, which allows us to capture nonlinearities in the relationship between the coefficient and the outcomes. Proportional odds assumptions of the final model were verified using plots of the mean of each candidate predictor across the levels of the ordinal outcome. We quantified the predictive accuracy of STRATIFY by calculating the discrimination using concordance (C statistic). We calculated the calibration by constructing a smooth nonparametric calibration curve of predicted versus observed outcome, which represents the bias in predicted values. We optimized the calibration curve at the low-risk end of the spectrum to minimize false negative results and define a threshold to accurately identify low-risk ED patients. We internally validated the calibration and discrimination for STRATIFY using bootstrap resampling to estimate the likely performance of the decision tool on a new patient sample from the same patient stream. All analyses were performed with R programming language.</p>

Describe whether and how the model was validated, either internally (e.g. bootstrapping, cross validation, random split sample) or externally (e.g. temporal validation, geographical validation, different setting, different type of participants): We internally validated the calibration and discrimination for STRATIFY using bootstrap resampling to estimate the likely performance of the decision tool on a new patient sample from the same patient stream (26). All analyses were performed with R programming language.			
Describe the performance measures of the model, e.g. (re)calibration, discrimination, (re)classification, net benefit: Discrimination: We used our methodology to filter 57 predictors through 8 principal components, which resulted in a decision tool with 13 variables readily obtained in the ED . Using the variable coefficients, we also developed a nomogram to assist the physician in estimating the risk of 30-day events. The C statistic of the STRATIFY decision tool was 0.68 (95% CI: 0.63 to 0.74). The C statistic of the tool in men was 0.64 (95% CI: 0.57 to 0.72), and for women it was 0.74 (95% CI: 0.67 to 0.81). When evaluating those patients with preserved ejection fraction ($\geq 45\%$), we found a C statistic of 0.72 (95% CI: 0.64 to 0.79). An elevated blood urea nitrogen and troponin level were found to be significant predictors of adverse events, whereas an elevated BNP, tachypnea, and use of dialysis trended toward a significant association with adverse events. Other variables included in the model were age, body mass index, diastolic blood pressure, sodium, oxygen saturation, QRS duration, and the use of supplemental oxygen or an angiotensin-converting enzyme inhibitor as an outpatient. These individual variables did not have a statistically significant association with the outcome in the final model. Figure 4 presents the calibration curve and shows that STRATIFY performed well in identifying patients whose risk of a 30-day adverse event was 10% or less. Looking at risk thresholds of 1%, 3%, 5%, and 10%, we found 0%, 1.4%, 13.0%, and 49.5% of patients, respectively, were considered to be low risk. Our test characteristics suggest the STRATIFY rule is highly sensitive for identifying patients at low risk of subsequent adverse events. Its negative predictive values are 100%, 96%, and 93% for identifying true low-risk patients at 3%, 5%, and 10% risk of death or subsequent adverse events. There were few deaths in those identified as being at low risk. There were no deaths in the cohort with $\leq 3\%$ risk of events, 1 death in the cohort with a 3% to 5% risk of events, and 6 deaths in the subjects with a 5% to 10% risk of events.			
Describe any participants who were excluded from the analysis: In the present study data were taken from patients with symptoms of AHF, 1853 were excluded because they refused to participate, 63 withdrew and 916 were not diagnosed with AHF.			
Describe missing data on predictors and outcomes as well as methods used for missing data: Then we evaluated descriptive statistics of candidate predictors for degree of missingness and level of information provided. A priori, predictors with missingness $>90\%$ were removed from consideration because their inclusion would minimize the usefulness of the tool. Categorical predictors with $>95\%$ prevalence for 1 level were also removed because of lack of variation and sufficient information to the tool. Descriptive statistics on the remaining potential predictors were calculated with median (interquartile range) or percentage (n), as appropriate. Missing data on remaining candidate predictors were imputed with single imputation that allowed for nonlinear transformations on the data to make optimal use of partial information recorded for each subject (26,28).			
	Dev	Val	
4.1 Were there a reasonable number of participants with the outcome?	Y		
4.2 Were continuous and categorical predictors handled appropriately?	Y		
4.3 Were all enrolled participants included in the analysis?	N		
4.4 Were participants with missing data handled appropriately?	Y		
4.5 Was selection of predictors based on univariable analysis avoided?	Y		
4.6 Were complexities in the data (e.g. censoring, competing risks, sampling of controls) accounted for appropriately?	Y		
4.7 Were relevant model performance measures evaluated appropriately?	Y		
4.8 Were model overfitting and optimism in model performance accounted for?	Y		
4.9 Do predictors and their assigned weights in the final model correspond to the results from the reported multivariable analysis?	Y		
Risk of bias introduced by the analysis	RISK: (low/ unclear)	high/ Low	

Rationale of bias rating: They show complete performance data, as well as missing data management.

Step 4: Overall assessment

Use the following tables to reach overall judgments about risk of bias and concerns for applicability of the prediction model evaluation (development and/or validation) across all assessed domains.

Reaching an overall judgment about risk of bias of the prediction model evaluation

Low risk of bias	If all domains were rated low risk of bias. If a <u>prediction model was developed without any external validation</u> , and it was rated as <u>low risk of bias for all domains</u> , consider downgrading to high risk of bias . Such a model can only be considered as low risk of bias, if the development was based on a very large data set and included some form of internal validation.
High risk of bias	If at least one domain is judged to be at high risk of bias .
Unclear risk of bias	If an unclear risk of bias was noted in at least one domain and it was low risk for all other domains.

Complete for each evaluation of a distinct model.

Reaching an overall judgment about applicability of the prediction model evaluation

Low concerns for applicability	If low concerns for applicability for all domains, the prediction model evaluation is judged to have low concerns for applicability .
High concerns for applicability	If high concerns for applicability for at least one domain, the prediction model evaluation is judged to have high concerns for applicability .
Unclear concerns for applicability	If unclear concerns (but no “high concern”) for applicability for at least one domain, the prediction model evaluation is judged to have unclear concerns for applicability overall.

Overall judgment about risk of bias and applicability of the prediction model evaluation

Overall judgment of risk of bias	RISK: (low/ high/ unclear)	
<i>Summary of sources of potential bias: Low</i>		
Overall judgment of applicability	CONCERN: (low/ high/ unclear)	
<i>Summary of applicability concerns: Low</i>		

PROBAST = Prediction model Risk Of Bias ASsessment Tool.

8. Delgado A, et al:

Step 2: Classify the type of prediction model evaluation

Classify the evaluation based on its aim			
Type of prediction study	PROBAST boxes to complete	Tick as appropriate	Definitions for type of prediction model study
Development only	Dev	X	Prediction model development without external validation. These studies may include internal validation methods such as bootstrapping and cross-validation techniques
Development and validation	Dev and Val		Prediction model development combined with external validation in other participants in the same article
Validation only	Val		External validation of existing (previously developed) model in other participants
<i>This table should be completed once for each publication being assessed and for each relevant outcome in your review.</i>			
Publication reference	Delgado A, Rodrigues B, Nunes S, Baptista R, Marmelo B, Moreira D. Acute Heart Failure Registry: Risk Assessment Model in Decompensated Heart Failure. Arq Bras Cardiol. 2016; 107(6):557-567		
Models of interest	Acute Heart Failure Registry: Risk Assessment Model in Decompensated Heart Failure		
Outcome of interest	Primary endpoint for score derivation was defined as all-cause mortality and / or rehospitalization for HF at 12 months. For score validation, the following endpoints were used: all-cause mortality and / or readmission for HF at 6, 12 and 24 months.		

Step 3: Assess risk of bias and applicability

DOMAIN 1: Participants			
A. Risk of Bias			
<i>Describe the sources of data and criteria for participant selection:.. All patients signed the informed consent form, according to the protocol. Inclusion criterion was diagnosis of acute HF, defined according to the European Society of Cardiology criteria.Exclusion criteria were high-output HF, high suspicion for acute coronary syndrome as the etiology of HF at hospital admission (including patients requiring urgent reperfusion therapy), acute myocarditis, infectious endocarditis, pulmonary infection, pulmonary arterial hypertension, and severe mitral stenosis. Patients admitted and discharged in the emergency service were also excluded.</i>			
		Dev	Val
1.1 Were appropriate data sources used, e.g. cohort, RCT or nested case-control study data?		Y	
1.2 Were all inclusions and exclusions of participants appropriate?		Y	
Risk of bias introduced by selection of participants	RISK: (low/unclear) high/	low	

<i>Rationale of bias rating: Applied correct criteria for inclusion and exclusion of participants</i>			
B. Applicability			
<i>Describe included participants, setting and dates: The total population consisted of 600 patients hospitalized for acute HF in a cardiology service of a non-tertiary hospital from 2009 to 2011. Of the 600 patients included, 95 were lost to clinical follow-up. In the population of 505 patients, six independent, predicting variables of the event (death/rehospitalization for HF) were identified using the endpoint in 242 patients. Then, 337 patients were classified according to the risk score as Group A (lower risk) or Group B (higher risk)</i>			
Concern that the included participants and setting do not match the review question	CONCERN: (low/ high/ unclear)	Low	
<i>Rationale of applicability rating: The classification of the participants in AHF was under the correct criteria, and they agree with the definition of the pathology in the research question.</i>			

DOMAIN 2: Predictors			
A. Risk of Bias			
<i>List and describe predictors included in the final model, e.g. definition and timing of assessment: Age ≥ 75 years, E/e' ratio ≥ 15, brain-type natriuretic peptide ≥ 400 pg/ML, Uremia ≥ 60 mg/dL, Natremia < 135 mEq/L, Without angiotensin-converting-enzyme inhibitor/ angiotensin II receptor blockers at discharge. The risk models used in acute HF have several particularities. First, patients are assessed at admission or at the emergency service, which generally prioritizes the assessment of a very short- term risk (during hospitalization) or a medium-term risk (from 2 to 6 months after discharge), as the identification of high risk patients contributes to a closer follow-up and more intensive therapy in this period when patients are more vulnerable. For this reason, the variables of these models may be similar but are slightly different as compared with those of chronic HF scores, in emphasizing easy, rapidly accessible clinical, demographic and analytical factors. Variables of anthropometry, clinical presentations, comorbidities, precipitating factors, echocardiographic measurements, intra-hospital treatment and medications prescribed at discharge were included. Data collection and electrocardiography were conducted at patient's admission in the emergency service.</i>			
		De v	Val
2.1 Were predictors defined and assessed in a similar way for all participants?		Y	
2.2 Were predictor assessments made without knowledge of outcome data?		Y	
2.3 Are all predictors available at the time the model is intended to be used?		Y	
Risk of bias introduced by predictors or their assessment	RISK: (low/ high/ unclear)	Low	
<i>Rationale of bias rating: It has a complete description of the list of selected predictor variables, as well as the way in which they were studied in patients.</i>			
B. Applicability			
Concern that the definition, assessment or timing of predictors in the model do not match the review question	CONCERN: (low/ high/ unclear)	High	
<i>Rationale of applicability rating: Not all variables were evaluated at the same time, so there may be changes in the performance of the study. In addition, the echocardiogram makes its use limited.</i>			
DOMAIN 3: Outcome			
A. Risk of Bias			

<p><i>Describe the outcome, how it was defined and determined, and the time interval between predictor assessment and outcome determination: Predict unfavorable prognostic events in patients hospitalized for acute heart failure syndromes and characterize a group at greater risk with respect to their clinical characteristics, treatment and results, during a 24-month follow-up (12 in referral cohort, 24 in validation cohort)</i></p>			
		Dev	Val
3.1 Was the outcome determined appropriately?		Y	
3.2 Was a pre-specified or standard outcome definition used?		Y	
3.3 Were predictors excluded from the outcome definition?		Y	
3.4 Was the outcome defined and determined in a similar way for all participants?		Y	
3.5 Was the outcome determined without knowledge of predictor information?		Y	
3.6 Was the time interval between predictor assessment and outcome determination appropriate?		Y	
Risk of bias introduced by the outcome or its determination		RISK: (low/ high/ unclear)	Low
<p><i>Rationale of bias rating: It has an adequate outcome definition that matches the research question in our review.</i></p>			
B. Applicability			
<p><i>At what time point was the outcome determined: It was determined during the follow-up in the future.</i></p> <p><i>If a composite outcome was used, describe the relative frequency/distribution of each contributing outcome:</i></p>			
Concern that the outcome, its definition, timing or determination do not match the review question		CONCERN: (low/ high/ unclear)	Low
<p><i>Rationale of applicability rating: The time in which the outcome is determined is consistent with the proposed research question.</i></p>			
DOMAIN 4: Analysis			
Risk of Bias			
<p><i>Describe numbers of participants, number of candidate predictors (for DEV only), outcome events and events per candidate predictor (for DEV only): Of the 600 patients included, 95 were lost to clinical follow-up. In the population of 505 patients, six independent, predicting variables of the event (death/rehospitalization for HF) were identified using the endpoint in 242 patients. Then, 337 patients were classified according to the risk score as Group A (lower risk) or Group B (higher risk). The present study does not present the total number of events or the candidate variables for predictors.</i></p>			
<p><i>Describe how the model was developed (predictor selection, optimism, risk groups, model performance): Continuous variables were reported as mean and standard deviation, and percentage of patients in the intervals obtained with the cutoff points. Categorical variables were described as absolute and relative frequencies (%). The Student's t-test was used for continuous variables (that had previously passed the Kolmogorov-Smirnov normality test) and the chi-square test for comparisons between categorical variables. Logistic regression analysis and Cox regression were performed when appropriate (95% confidence interval). A significance level of p<0.05 was adopted. Of the 600 patients included, 95 were lost to clinical follow-up. In the population of 505 patients, six independent, predicting variables of the event (death/rehospitalization for HF) were identified using the endpoint in 242 patients. Then, 337 patients were classified according to the risk score as Group A (lower risk) or Group B (higher risk). It is important to assess this prognostic score regarding its discrimination and calibration. Discrimination was estimated by the area under the curve (AUC), and calibration was estimated by the Hosmer-Lemeshow test.</i></p>			

Describe whether and how the model was validated, either internally (e.g. bootstrapping, cross validation, random split sample) or externally (e.g. temporal validation, geographical validation, different setting, different type of participants): they don't mention the validation method		
Describe the performance measures of the model, e.g. (re)calibration, discrimination, (re)classification, net benefit: Discrimination: Therefore, the area under the ROC curve for the endpoint (mortality and/or rehospitalization at 12 months) was 0.74, with an intermediate discrimination score. The score was predictor of the event ($p < 0.001$), with 65% accuracy. The article mentions that they used hosmer lemeshow for calibration, however neither the graph nor the results are commented on in the article.		
Describe any participants who were excluded from the analysis: Of the 600 patients included, 95 were lost to clinical follow-up.		
Describe missing data on predictors and outcomes as well as methods used for missing data: do not describe the handling of missing data		
	Dev	Val
4.1 Were there a reasonable number of participants with the outcome?	N	
4.2 Were continuous and categorical predictors handled appropriately?	Y	
4.3 Were all enrolled participants included in the analysis?	N	
4.4 Were participants with missing data handled appropriately?	NI	
4.5 Was selection of predictors based on univariable analysis avoided?	PY	
4.6 Were complexities in the data (e.g. censoring, competing risks, sampling of controls) accounted for appropriately?	PN	
4.7 Were relevant model performance measures evaluated appropriately?	PY	
4.8 Were model overfitting and optimism in model performance accounted for?	NI	
4.9 Do predictors and their assigned weights in the final model correspond to the results from the reported multivariable analysis?	Y	
Risk of bias introduced by the analysis	RISK: (low/ high/ unclear)	High
Rationale of bias rating: They don't show data on calibration, or the handling of missing information. There are also gaps in the analysis, since the candidate variables and the exact number of events in the article are not described.		

Step 4: Overall assessment

Use the following tables to reach overall judgments about risk of bias and concerns for applicability of the prediction model evaluation (development and/or validation) across all assessed domains.

Reaching an overall judgment about risk of bias of the prediction model evaluation	
Low risk of bias	If all domains were rated low risk of bias. If a prediction model was developed without any external validation, and it was rated as <u>low risk of bias for all domains</u> , consider downgrading to high risk of bias . Such a model can only be considered as low risk of bias, if the development was based on a very large data set and included some form of internal validation.
High risk of bias	If at least one domain is judged to be at high risk of bias .
Unclear risk of bias	If an unclear risk of bias was noted in at least one domain and it was low risk for all other domains.

Complete for each evaluation of a distinct model.

Reaching an overall judgment about applicability of the prediction model evaluation	
Low concerns for applicability	If low concerns for applicability for all domains, the prediction model evaluation is judged to have low concerns for applicability .
High concerns for applicability	If high concerns for applicability for at least one domain, the prediction model evaluation is judged to have high concerns for applicability .
Unclear concerns for applicability	If unclear concerns (but no “high concern”) for applicability for at least one domain, the prediction model evaluation is judged to have unclear concerns for applicability overall.

Overall judgment about risk of bias and applicability of the prediction model evaluation		
Overall judgment of risk of bias	RISK: (low/ high/ unclear)	
Summary of sources of potential bias: high		
Overall judgment of applicability	CONCERN: (low/ high/ unclear)	
Summary of applicability concerns: high		

PROBAST = Prediction model Risk Of Bias ASsessment Tool.

9. DeVore AD, et al

Step 2: Classify the type of prediction model evaluation

Classify the evaluation based on its aim

Type of prediction study	PROBAST boxes to complete	Tick as appropriate	Definitions for type of prediction model study
Development only	Dev		Prediction model development without external validation. These studies may include internal validation methods such as bootstrapping and cross-validation techniques
Development and validation	Dev and Val	X	Prediction model development combined with external validation in other participants in the same article
Validation only	Val		External validation of existing (previously developed) model in other participants
<i>This table should be completed once for each publication being assessed and for each relevant outcome in your review.</i>			
Publication reference	DeVore AD, Greiner MA, Sharma PP, Qualls LG, Schulte PJ, Cooper LB, et al. Development and validation of a risk model for in-hospital worsening heart failure from the Acute Decompensated Heart Failure National Registry (ADHERE). Am Heart J 2016; 178:198-205.		
Models of interest	Development and validation of a risk model for in-hospital worsening heart failure from the Acute Decompensated Heart Failure National Registry (ADHERE)		
Outcome of interest	Develop and validate a risk model for in-hospital worsening heart failure. Our definition of in-hospital worsening heart failure included events such as escalation of medical therapy (eg, inotropic medications) >12 hours after admission.		

Step 3: Assess risk of bias and applicability

DOMAIN 1: Participants			
A. Risk of Bias			
<p><i>Describe the sources of data and criteria for participant selection: We used data from ADHERE and fee-for-service Medicare standard analytic claim files for this analysis, similar to previous analyses. ADHERE was a large, multicenter registry of patients hospitalized with acute heart failure at >300 community and academic medical centers throughout the United States. Data collected on >185,000 patients between January 2001 and March 2006 included demographic characteristics, comorbid conditions, medications, laboratory test results, procedures, and in-hospital outcomes. We obtained standard analytic files for fee-for-service Medicare beneficiaries from the US Centers for Medicare & Medicaid Services. We obtained information on Medicare eligibility and enrollment from the denominator files and hospital service dates and diagnosis codes from the inpatient files. The carrier files contain claims from noninstitutional professional providers and include Healthcare Common Procedure Coding System codes and service dates, which we used to verify in-hospital procedures. We considered patients to have a complicated hospital presentation if they met any of the following criteria: initiated inotropic medications or an intravenous vasodilator within 12 hours of hospital presentation, were admitted initially to the intensive care unit, or received advanced medical therapy (ie, mechanical circulatory support, mechanical ventilation, or hemodialysis) on the first inpatient day.</i></p>			
		Dev	Val
1.1 Were appropriate data sources used, e.g. cohort, RCT or nested case-control study data?		Y	Y
1.2 Were all inclusions and exclusions of participants appropriate?		Y	Y
Risk of bias introduced by selection of participants	RISK: (low/ unclear) high/	low	Low
<i>Rationale of bias rating: Applied correct criteria for inclusion of participants</i>			

B. Applicability			
<p><i>Describe included participants, setting and dates:</i> There were 15,640 (66%) patients in the derivation sample and 8,056 (34%) in the validation sample. In the derivation sample, the mean age was 81 years (SD, 7.7 years); mean systolic blood pressure was 145 mm Hg (SD, 29.8 mm Hg); 70% of patients had a prior heart failure diagnosis, with 12% hospitalized for heart failure in the previous 6 months; 34% had left ventricular ejection fraction <40%; mean BNP was 1067 pg/mL (SD, 1012 pg/mL); and mean creatinine was 1.5 mg/dL (SD, 0.9 mg/dL). Baseline characteristics were well matched in the derivation and validation samples. In-hospital worsening heart failure occurred in 2,403 (15.4%) patients in the derivation sample and 1,257 (15.6%) in the validation sample (P = .63).</p>			
Concern that the included participants and setting do not match the review question	CONCERN: (low/ high/ unclear)	Low	Low
<p><i>Rationale of applicability rating:</i> The characteristics of the study population are consistent with the criteria of the research question.</p>			
A. Risk of Bias			
<p><i>List and describe predictors included in the final model, e.g. definition and timing of assessment:</i> After testing candidate variables in preliminary models, the final reduced model included the following admission variables: age, pulse, systolic blood pressure, left ventricular ejection fraction, BNP level, troponin (positive or negative), sodium, blood urea nitrogen, and creatinine. The ADHERE risk model uses 9 clinical variables routinely assessed at the time of admission and provides good discrimination for use inpatients hospitalized with acute heart failure to identify those at risk for in-hospital worsening heart failure.</p>			
		De v	Val
2.1 Were predictors defined and assessed in a similar way for all participants?		Y	Y
2.2 Were predictor assessments made without knowledge of outcome data?		Y	Y
2.3 Are all predictors available at the time the model is intended to be used?		Y	Y
Risk of bias introduced by predictors or their assessment	RISK: (low/ high/ unclear)	Low	Low
<p><i>Rationale of bias rating:</i> It has a complete description of the list of selected predictor variables, as well as the way in which they were studied in patients.</p>			
B. Applicability			
Concern that the definition, assessment or timing of predictors in the model do not match the review question	CONCERN: (low/ high/ unclear)	Low	Low
<p><i>Rationale of applicability rating:</i></p> <p>The evaluation time of the predictive variables and the applicability of these are consistent with the research question of the recent systematic review.</p>			
DOMAIN 3: Outcome			
A. Risk of Bias			
<p><i>Describe the outcome, how it was defined and determined, and the time interval between predictor assessment and outcome determination:</i> Validate a risk model for in-hospital worsening heart failure. Our definition of in-hospital worsening heart failure included events such as escalation of medical therapy (eg, inotropic medications) >12 hours after admission.</p>			
		Dev	Val
3.1 Was the outcome determined appropriately?		Y	Y
3.2 Was a pre-specified or standard outcome definition used?		Y	Y
3.3 Were predictors excluded from the outcome definition?		Y	Y
3.4 Was the outcome defined and determined in a similar way for all participants?		Y	Y

3.5 Was the outcome determined without knowledge of predictor information?	Y	Y
3.6 Was the time interval between predictor assessment and outcome determination appropriate?	Y	Y
Risk of bias introduced by the outcome or its determination	RISK: (low/ high/ unclear)	Low Low
<i>Rationale of bias rating: It has an adequate outcome definition that matches the research question in our review.</i>		
B. Applicability		
<p><i>At what time point was the outcome determined:</i> <i>It was determined during the follow-up in the future, during the time of hospitalization.</i></p> <p><i>If a composite outcome was used, describe the relative frequency/distribution of each contributing outcome:</i></p>		
Concern that the outcome, its definition, timing or determination do not match the review question	CONCERN: (low/ high/ unclear)	Low Low
<i>Rationale of applicability rating: The time in which the outcome is determined is consistent with the proposed research question.</i>		
DOMAIN 4: Analysis		
Risk of Bias		
<p><i>Describe numbers of participants, number of candidate predictors (for DEV only), outcome events and events per candidate predictor (for DEV only): There were 15,640 (66%) patients in the derivation sample and 8,056 (34%) in the validation sample. We carefully selected candidate risk prediction variables based on clinical judgment and previously published models, and we adhered to the principle of 10 events per examined variable to minimize prediction error and avoid overfitting.</i></p>		
<p><i>Describe how the model was developed (predictor selection, optimism, risk groups, model performance): We considered variables that were routinely assessed on admission including demographic characteristics, medical history, biomarkers, and renal function. We used logistic regression models to predict in-hospital worsening heart failure. We incorporated generalized estimating equations to account for hospital clustering. We developed the prediction model in the derivation sample and then applied the results from the model in the validation sample. We examined quadratic variables and linear splines when continuous variables failed the Box-Tidwell linearity test. To further refine the model for routine clinical use, we examined preliminary model results and developed a reduced model based on clinical judgment and the least absolute shrinkage and selection operator reduction method. We categorized patients into 2 groups according to their experience during the index hospitalization: patients with in-hospital worsening heart failure and patients without in-hospital worsening heart failure. As in previous prior analyses, we classified patients as having in-hospital worsening heart failure if they met any of the following criteria: initiated inotropic medications or an intravenous vasodilator >12 hours after hospital presentation, were transferred to the intensive care unit, or received advanced medical therapy (ie, mechanical circulatory support, mechanical ventilation, or hemodialysis) after the first inpatient day. We classified the remaining patients as having no in-hospital worsening heart failure.</i></p>		
<p><i>Describe whether and how the model was validated, either internally (e.g. bootstrapping, cross validation, random split sample) or externally (e.g. temporal validation, geographical validation, different setting, different type of participants): Using simple random selection, we constructed a derivation sample consisting of 66%(internal validation) of the study cohort and a validation sample consisting of the remaining 34% of the study cohort. we externally evaluated the risk model using data from the Acute Study of Nesiritide in Decompensated Heart Failure (ASCEND-HF).</i></p>		
<p><i>Describe the performance measures of the model, e.g. (re)calibration, discrimination, (re)classification, net benefit: Discrimination: In both the derivation and validation samples, the model was well calibrated for predicting in-hospital worsening heart failure. The Eavg was 0.43 for the derivation sample and 0.14 for the validation sample, indicating overall consistency between observed and predicted outcome rates. The</i></p>		

<p>percentages of predicted and observed in-hospital worsening heart failure were similar within deciles (Figure) (validation Hosmer-Lemeshow statistic, 8.71; $P = .14$). The model had good discrimination in the derivation (c statistic, 0.74) and validation samples (c statistic, 0.72) (Table III). Using the 15% threshold for predicted probabilities in the validation sample, 37% of the participants who actually had in-hospital worsening heart failure were predicted to have in-hospital worsening heart failure. The sensitivity of the model improved as the threshold became less stringent, reaching 75% at the 50% threshold. On the other hand, using the 15% threshold, 89% of the participants who actually had no in-hospital worsening heart failure were predicted to have no in-hospital worsening heart failure. The specificity of the model declined as the threshold became less stringent. For external validation, there were 1,464 of 7,141 (21%) patients who met the study criteria in ASCEND-HF. Using the definition of worsening heart failure from the ASCEND-HF protocol, there were 65 (4.4%) events. The risk model was poorly calibrated as evidenced by the absolute difference between the percentages of observed and predicted worsening heart failure (E_{avg}, 14.8) and Hosmer-Lemeshow test (statistic, 244.1; $P < .001$). There was modest discrimination with a c statistic of 0.63.</p>			
<p><i>Describe any participants who were excluded from the analysis:</i> we excluded patients with a complicated hospital presentation. We excluded patients without B-type natriuretic peptide (BNP) results because BNP was as a key candidate variable for the risk model. We also excluded patients who were admitted on an elective basis for the index hospitalization and those who did not have fee-for-service Medicare coverage for at least 6 months before the index hospitalization.</p>			
<p><i>Describe missing data on predictors and outcomes as well as methods used for missing data:</i> As in previous ADHERE analyses, a positive troponin test result was defined as a cardiac troponin I level of ≥ 1.0 ng/mL or a cardiac troponin T level of ≥ 0.1 ng/mL.²⁰ A qualitative troponin test indicator (ie, positive or negative) was used when the quantitative value was missing. We used Medicare claims from 6 months before the index admission to identify previous heart failure hospitalizations (International Classification of Diseases, Ninth Revision, Clinical Modification, primary diagnosis code 428.x, 402.x1, 404.x1, or 404.x3). For variables with low rates of missingness (ie, $\leq 5\%$ of records), we imputed missing continuous variables to the overall median value, dichotomous variables to “no,” and multichotomous variables to the most frequent categorical value.^{21,22} For variables with $\geq 5\%$ missingness (including smoking status, troponin, and ejection fraction), we treated missing values as a separate category.</p>			
		Dev	Val
4.1 Were there a reasonable number of participants with the outcome?		Y	Y
4.2 Were continuous and categorical predictors handled appropriately?		Y	Y
4.3 Were all enrolled participants included in the analysis?		Y	Y
4.4 Were participants with missing data handled appropriately?		Y	Y
4.5 Was selection of predictors based on univariable analysis avoided?		Y	Y
4.6 Were complexities in the data (e.g. censoring, competing risks, sampling of controls) accounted for appropriately?		Y	Y
4.7 Were relevant model performance measures evaluated appropriately?		Y	Y
4.8 Were model overfitting and optimism in model performance accounted for?		Y	Y
4.9 Do predictors and their assigned weights in the final model correspond to the results from the reported multivariable analysis?		Y	Y
Risk of bias introduced by the analysis	RISK: (low/ high/ unclear)	Low	Low
<p><i>Rationale of bias rating:</i> The study provides clear data on the analysis, as well as the management of lost data and good performance.</p>			

Step 4: Overall assessment

Use the following tables to reach overall judgments about risk of bias and concerns for applicability of the prediction model evaluation (development and/or validation) across all assessed domains.

Reaching an overall judgment about risk of bias of the prediction model evaluation	
Low risk of bias	If all domains were rated low risk of bias. If a prediction model was developed without any external validation, and it was rated as <u>low risk of bias for all domains</u> , consider downgrading to high risk of bias . Such a model can only be considered as low risk of bias, if the development was based on a very large data set and included some form of internal validation.
High risk of bias	If at least one domain is judged to be at high risk of bias .
Unclear risk of bias	If an unclear risk of bias was noted in at least one domain and it was low risk for all other domains.

Complete for each evaluation of a distinct model.

Reaching an overall judgment about applicability of the prediction model evaluation	
Low concerns for applicability	If low concerns for applicability for all domains, the prediction model evaluation is judged to have low concerns for applicability .
High concerns for applicability	If high concerns for applicability for at least one domain, the prediction model evaluation is judged to have high concerns for applicability .
Unclear concerns for applicability	If unclear concerns (but no “high concern”) for applicability for at least one domain, the prediction model evaluation is judged to have unclear concerns for applicability overall.

Overall judgment about risk of bias and applicability of the prediction model evaluation		
Overall judgment of risk of bias	RISK: (low/ high/ unclear)	
Summary of sources of potential bias: Low		
Overall judgment of applicability	CONCERN: (low/ high/ unclear)	
Summary of applicability concerns: Low		

PROBAST = Prediction model Risk Of Bias ASsessment Tool.

10. Diercks DR, et al:

Step 2: Classify the type of prediction model evaluation

Classify the evaluation based on its aim			
Type of prediction study	PROBAST boxes to complete	Tick as appropriate	Definitions for type of prediction model study
Development only	Dev		Prediction model development without external validation. These studies may include internal validation methods such as bootstrapping and cross-validation techniques
Development and validation	Dev and Val		Prediction model development combined with external validation in other participants in the same article
Validation only	Val	X	External validation of existing (previously developed) model in other participants
<i>This table should be completed once for each publication being assessed and for each relevant outcome in your review.</i>			
Publication reference	Diercks DB, Fonarow GC, Kirk JD, Emerman CL, Hollander JE, Weber JE. Risk Stratification in Women Enrolled in the Acute Decompensated Heart Failure National Registry Emergency Module (ADHERE-EM). ACADEMIC EMERGENCY MEDICINE 2008; 15(2):151–158		
Models of interest	Development and validation of a risk model for in-hospital worsening heart failure from the Acute Decompensated Heart Failure National Registry (ADHERE)		
Outcome of interest	Ability of a published risk stratification model to predict outcomes in women.		

Step 3: Assess risk of bias and applicability

DOMAIN 1: Participants			
A. Risk of Bias			
<p><i>Describe the sources of data and criteria for participant selection: We used the Acute Decompensated Heart Failure National Registry Emergency Module (ADHERE-EM) database to evaluate gender differences in the presentation, treatment, and outcome of patients with ADHF in the emergency department (ED). The ADHERE-EM database was created as part of the ADHERE registry program, and patients were enrolled between January 2004 and September 2005. ADHERE is a large, national, multicenter registry designed to collect data on the demographics, clinical characteristics, treatment patterns, and outcomes of patients hospitalized for ADHF.³⁰ The data were collected at 83 hospitals, including 53 academic centers. The goal of ADHERE-EM was to specifically characterize the initial ED evaluation, treatment, and subsequent inpatient management of patients presenting to the ED with ADHF. Therefore, the ADHERE-EM contains detailed data regarding the initial ED presentation, treatment, and disposition not contained in the ADHERE registry. The registry entries consisted of all consecutive adult patients with either an ED or a hospital discharge diagnosis of ADHF, ADHF that was determined to be clinically present by the patient care team, or ADHF as documented in the diagnosis-related group (DRG) code or ambulatory procedure code. The diagnosis of ADHF was based on the DRG in admitted patients. ADHF had to be the focus of treatment for the patient to be included. Importantly, registry participation did not require any alteration of treatment or ED care, and entry of data into the registry was not contingent on the use of any particular therapeutic agent or treatment regimen.</i></p>			
		Dev	Val
1.1 Were appropriate data sources used, e.g. cohort, RCT or nested case-control study data?			Y
1.2 Were all inclusions and exclusions of participants appropriate?			Y
Risk of bias introduced by selection of participants	RISK: (low/unclear) high/		Low

<i>Rationale of bias rating: Applied correct criteria for inclusion of participants</i>			
B. Applicability			
<i>Describe included participants, setting and dates:</i> A total of 10,984 patient records were available in the ADHERE-EM database. Of these, 5,248 (47.8%) were men and 5,736 (52.2%) were women. Heart failure was the primary discharge diagnosis in 97%, and 83% had heart failure as the ED admitting diagnosis. The mean age of the overall patient population was 73.1 years, and men were younger than women (Table 1). In the overall patient population, 76% had a prior history of heart failure, 45% had a history of diabetes, and 79% had a history of hypertension. The ADHERE-EM database was created as part of the ADHERE registry program, and patients were enrolled between January 2004 and September 2005.			
Concern that the included participants and setting do not match the review question	CONCERN: (low/ high/ unclear)		Low
<i>Rationale of applicability rating: The characteristics of the study population are consistent with the criteria of the research question.</i>			
DOMAIN 2: Predictors			
A. Risk of Bias			
<i>List and describe predictors included in the final model, e.g. definition and timing of assessment:</i> age, pulse, systolic blood pressure, left ventricular ejection fraction, BNP level, troponin (positive or negative), sodium, blood urea nitrogen, and creatinine. The ADHERE risk model uses 9 clinical variables routinely assessed at the time of admission and provides good discrimination for use inpatients hospitalized with acute heart failure to identify those at risk for in-hospital worsening heart failure.			
		De v	Val
2.1 Were predictors defined and assessed in a similar way for all participants?			Y
2.2 Were predictor assessments made without knowledge of outcome data?			Y
2.3 Are all predictors available at the time the model is intended to be used?			Y
Risk of bias introduced by predictors or their assessment	RISK: (low/ high/ unclear)		Low
<i>Rationale of bias rating: It has a complete description of the list of selected predictor variables, as well as the way in which they were studied in patients.</i>			
B. Applicability			
Concern that the definition, assessment or timing of predictors in the model do not match the review question	CONCERN: (low/ high/ unclear)		Low
<i>Rationale of applicability rating:</i>			
<i>The evaluation time of the predictive variables and the applicability of these are consistent with the research question of the recent systematic review.</i>			
DOMAIN 3: Outcome			
A. Risk of Bias			
<i>Describe the outcome, how it was defined and determined, and the time interval between predictor assessment and outcome determination:</i> This was a retrospective chart review to examine the ability of a published risk-stratification model to predict outcomes in women.			
		Dev	Val
3.1 Was the outcome determined appropriately?			Y
3.2 Was a pre-specified or standard outcome definition used?			Y
3.3 Were predictors excluded from the outcome definition?			Y
3.4 Was the outcome defined and determined in a similar way for all participants?			Y
3.5 Was the outcome determined without knowledge of predictor information?			Y

3.6 Was the time interval between predictor assessment and outcome determination appropriate?			Y
Risk of bias introduced by the outcome or its determination	RISK: (low/ high/ unclear)		Low
<i>Rationale of bias rating: It has an adequate outcome definition that matches the research question in our review.</i>			
B. Applicability			
<p><i>At what time point was the outcome determined:</i> <i>It was determined during the follow-up in the future, during the time of hospitalization.</i></p> <p><i>If a composite outcome was used, describe the relative frequency/distribution of each contributing outcome:</i></p>			
Concern that the outcome, its definition, timing or determination do not match the review question	CONCERN: (low/ high/ unclear)		Low
<i>Rationale of applicability rating: The time in which the outcome is determined is consistent with the proposed research question.</i>			

DOMAIN 4: Analysis		
Risk of Bias		
<p><i>Describe numbers of participants, number of candidate predictors (for DEV only), outcome events and events per candidate predictor (for DEV only):</i> A total of 10,984 patient records were available in the ADHERE-EM database. Of these, 5,248 (47.8%) were men and 5,736 (52.2%) were women. Heart failure was the primary discharge diagnosis in 97%, and 83% had heart failure as the ED admitting diagnosis. The mean age of the overall patient population was 73.1 years, and men were younger than women (Table 1). In the overall patient population, 76% had a prior history of heart failure, 45% had a history of diabetes, and 79% had a history of hypertension. A slightly higher proportion of men had a history of heart failure or diabetes, and more women had a history of hypertension. Preserved systolic function was present in a significantly greater proportion of women compared to men.</p>		
<p><i>Describe how the model was developed (predictor selection, optimism, risk groups, model performance):</i> Descriptive statistics were used to describe gender differences in presentation, demographics, treatments, and laboratory values. A logistic regression model adjusting for baseline covariates that have previously been shown in the main ADHERE registry to be associated with mortality (blood urea nitrogen [BUN], systolic blood pressure [sBP], serum creatinine [SCr], age, and dyspnea at rest) was performed. For the risk stratification analysis, patients were categorized according to the five risk groups identified by the previously derived and validated classification and regression tree (CART) analysis of the main ADHERE registry: high risk (BUN $\geq$ 43 mg/dL, sBP $<$ 115 mm Hg, and SCr $\geq$ 2.75 mg/dL); intermediate risk 1 (BUN $\geq$ 43 mg/dL, sBP $<$ 115 mm Hg, and SCr $<$ 2.75 mg/dL); intermediate risk 2 (BUN $\geq$ 43 mg/dL and sBP $\geq$ 115 mm Hg); intermediate risk 3 (BUN $<$ 43 mg/dL and sBP $<$ 115 mm Hg); and low risk (BUN $<$ 43 mg/dL and sBP $\geq$ 115 mm Hg).²⁴ A complete case analysis was performed. The predictive value of this model was then assessed by determination of mortality odds ratios (ORs) and 95% confidence intervals (CIs) between risk groups for men and women. These analyses were performed using Version 8.2 of SAS software (SAS Institute, Inc., Cary, NC).</p>		
<p><i>Describe whether and how the model was validated, either internally (e.g. bootstrapping, cross validation, random split sample) or externally (e.g. temporal validation, geographical validation, different setting, different type of participants):</i> It is validated externally when using a different population from the original study</p>		
<p><i>Describe the performance measures of the model, e.g. (re)calibration, discrimination, (re)classification, net benefit:</i> In-hospital outcomes and procedures were similar among men and women. In-hospital mortality was 3% for men and 2.8% for women (OR 1.04, 95% CI = 0.83 to 1.30). The risk of death was similar between men and women after adjusting for covariates (OR 1.10, 95% CI = 0.86 to 1.40). No gender differences were present in other important in-hospital clinical procedures, including dialysis or need for mechanical ventilation. The decision tree generated by CART analysis categorized men and women into high-, intermediate-, and low-risk groups. In the study cohort, 98.5% of all subjects had complete data and were included in the analysis. The clinical characteristics of the five groups by gender are shown. Patients in the higher risk groups tended to be older and diabetic and have systolic dysfunction compared to patients in the lower risk groups. Differences were observed in the proportion of men and women at each risk node. Women were more likely to be in the lowest risk node (OR 1.3, 95% CI = 1.2 to 1.4), while men were more likely to be in the highest risk node (OR 1.7, 95% CI = 1.2 to 2.4). No significant difference in the rate of mortality by gender was detected within risk categories. The risk categorization method stratified in-hospital mortality risk regardless of gender. The mortality OR between the high- and low-risk groups was 9.3 (95% CI = 4.2 to 20.3, $p < 0.0001$) in women and 8.6 (95% CI = 4.0 to 18.1, $p < 0.0001$) in men. Discrimination between the intermediate-risk and the high-risk group was not as robust.</p>		
<p><i>Describe any participants who were excluded from the analysis:</i> The patients who were excluded are not described, since the data obtained was from a selection from a database.</p>		
<p><i>Describe missing data on predictors and outcomes as well as methods used for missing data:</i> Doesn't describe what was done with the lost data</p>		
	Dev	Val
4.1 Were there a reasonable number of participants with the outcome?		Y
4.2 Were continuous and categorical predictors handled appropriately?		Y

4.3 Were all enrolled participants included in the analysis?		Y
4.4 Were participants with missing data handled appropriately?		PY
4.5 Was selection of predictors based on univariable analysis avoided?		Y
4.6 Were complexities in the data (e.g. censoring, competing risks, sampling of controls) accounted for appropriately?		N
4.7 Were relevant model performance measures evaluated appropriately?		N
4.8 Were model overfitting and optimism in model performance accounted for?		Y
4.9 Do predictors and their assigned weights in the final model correspond to the results from the reported multivariable analysis?		Y
Risk of bias introduced by the analysis	RISK: (low/ high/ unclear)	High
<i>Rationale of bias rating: . It has an adequate performance evaluation, although it does not describe the lost data, this is probably because the selection of participants was from a database.</i>		

Step 4: Overall assessment

Use the following tables to reach overall judgments about risk of bias and concerns for applicability of the prediction model evaluation (development and/or validation) across all assessed domains.

Reaching an overall judgment about risk of bias of the prediction model evaluation	
Low risk of bias	If all domains were rated low risk of bias. If a <u>prediction model was developed without any external validation</u> , and it was rated as <u>low risk of bias for all domains</u> , consider downgrading to high risk of bias . Such a model can only be considered as low risk of bias, if the development was based on a very large data set and included some form of internal validation.
High risk of bias	If at least one domain is judged to be at high risk of bias .
Unclear risk of bias	If an unclear risk of bias was noted in at least one domain and it was low risk for all other domains.

Complete for each evaluation of a distinct model.

Reaching an overall judgment about applicability of the prediction model evaluation	
Low concerns for applicability	If low concerns for applicability for all domains, the prediction model evaluation is judged to have low concerns for applicability .
High concerns for applicability	If high concerns for applicability for at least one domain, the prediction model evaluation is judged to have high concerns for applicability .
Unclear concerns for applicability	If unclear concerns (but no “high concern”) for applicability for at least one domain, the prediction model evaluation is judged to have unclear concerns for applicability overall.

Overall judgment about risk of bias and applicability of the prediction model evaluation		
Overall judgment of risk of bias	RISK: (low/ high/	

	unclear)	
Summary of sources of potential bias: High		
Overall judgment of applicability	CONCERN: (low/ high/ unclear)	
Summary of applicability concerns: Low		

PROBAST = Prediction model Risk Of Bias ASsessment Tool.

11. Eurlings LW, et al:

Step 2: Classify the type of prediction model evaluation

Classify the evaluation based on its aim			
Type of prediction study	PROBAST boxes to complete	Tick as appropriate	Definitions for type of prediction model study
Development only	Dev		Prediction model development without external validation. These studies may include internal validation methods such as bootstrapping and cross-validation techniques
Development and validation	Dev and Val		Prediction model development combined with external validation in other participants in the same article
Validation only	Val	X	External validation of existing (previously developed) model in other participants
<i>This table should be completed once for each publication being assessed and for each relevant outcome in your review.</i>			
Publication reference	Eurlings LW, Sanders-Van Wijk S, Van Kraaij DJ, Van Kimmenade R, Meeder JG, Kamp O, et al. Risk Stratification With the Use of Serial N-Terminal Pro-B-Type Natriuretic Peptide Measurements During Admission and Early After Discharge in Heart Failure Patients: Post Hoc Analysis of the PRIMA Study. Journal of Cardiac Failure. 2014. 20(12): 881-890.		
Models of	NT-proBNP score		

interest	
Outcome of interest	Primary outcome measures were mortality and the combined end point of HF readmission or mortality within the follow-up period after the outpatient visit 1 month after discharge. Secondary end points encompassed the primary end points reached at 90, 180, and 365 days of follow-up.

Step 3: Assess risk of bias and applicability

DOMAIN 1: Participants			
A. Risk of Bias			
<p><i>Describe the sources of data and criteria for participant selection: This was a post hoc analysis of patients included in the PRIMA (Can P_{RO}-brain-natriuretic peptide guided therapy of chronic heart failure I_Mprove heart fAilure morbidity and mortality?) study, a prospective randomized multicenter study assessing the effect of management of chronic HF guided by individual NTproBNP targets. Inclusion criteria have been published previously. In short, patients were included during hospital admission for acute HF. NT-proBNP concentration at admission was required to be $\geq 1,700$ pg/mL, and included patients also needed to demonstrate a decrease in NT-proBNP concentration of at least 10% with a minimum of 850 pg/mL during admission. At discharge, patients were randomized to outpatient treatment that was either clinically guided where NT-proBNP was measured but not revealed to the physician, or to outpatient treatment where NT-proBNP levels were provided to guide therapy. The follow-up period was up to 2 years. For the present analysis, patients with outpatient NTproBNP concentration available 1 month after hospital discharge were included. As a result, patients not attending the outpatient clinic 1 month after discharge (because of death, readmission, or any other reason) were not included. All events occurring before the outpatient visit 1 month after hospital discharge were censored.</i></p>			
		De v	Val
1.1 Were appropriate data sources used, e.g. cohort, RCT or nested case-control study data?		Y	
1.2 Were all inclusions and exclusions of participants appropriate?		Y	
Risk of bias introduced by selection of participants	RISK: (low/ high/ unclear)	Low	
<i>Rationale of bias rating: Applied correct criteria for inclusion of participants</i>			
B. Applicability			
<p><i>Describe included participants, setting and dates: In 309 out of 345 patients included in the PRIMA study, NT-pro BNP levels at admission, discharge, and 1 month after hospital discharge were available. Patient characteristics at baseline and 1 month after hospital discharge are presented in. Patients were overall elderly and predominantly male, more than one-half had coronary artery disease, and about one-half had an ischemic etiology of HF. At admission because of acute HF, median NTproBNP concentration was clearly elevated, and during admission the median decrease in NT-proBNP concentration was >60%. One month after hospital discharge, median NT-proBNP concentration was 2,538 pg/mL.</i></p>			
Concern that the included participants and setting do not match the review question	CONCERN: (low/ high/ unclear)	Low	
<i>Rationale of applicability rating: The characteristics of the study population are consistent with the criteria of the research question.</i>			

DOMAIN 2: Predictors	
A. Risk of Bias	
<p><i>List and describe predictors included in the final model, e.g. definition and timing of assessment:</i></p> <p><i>Mortality</i></p> <p><i>Clinical risk factors</i></p> <p>+ Inpatient change in NT-proBNP</p> <p>+ Early outpatient change in NT-proBNP</p>	

+ NT-proBNP at 1 month follow-up Clinical risk factors + inpatient change in NT-proBNP + Early outpatient change NT-proBNP + NT-proBNP at 1 month Clinical risk factors + early outpatient change in NT-proBNP + Inpatient change in NT-proBNP + NT-proBNP at 1 month Clinical risk factors + NT-proBNP at 1 month + Inpatient change in NT-proBNP + early outpatient change in NT-proBNP HF-related readmission or mortality Clinical risk factors + Inpatient change in NT-proBNP + Early outpatient change in NT-proBNP + NT-proBNP at 1 month follow-up Clinical risk factors + inpatient change in NT-proBNP + Early outpatient change in NT-proBNP + NT-proBNP at 1 month Clinical risk factors + early outpatient change NT-proBNP + Inpatient change in NT-proBNP + NT-proBNP at 1 month Clinical risk factors + NT-proBNP at 1 month + Inpatient change in NT-proBNP + early outpatient change in NT-proBNP			
	De v	Val	
2.1 Were predictors defined and assessed in a similar way for all participants?	Y		
2.2 Were predictor assessments made without knowledge of outcome data?	Y		
2.3 Are all predictors available at the time the model is intended to be used?	Y		
Risk of bias introduced by predictors or their assessment	RISK: (low/ high/ unclear)	Low	
<i>Rationale of bias rating: It has a complete description of the list of selected predictor variables, as well as the way in which they were studied in patients.</i>			
B. Applicability			
Concern that the definition, assessment or timing of predictors in the model do not match the review question	CONCERN: (low/ high/ unclear)	Low	
<i>Rationale of applicability rating: The evaluation time of the predictive variables and the applicability of these are consistent with the research question of the recent systematic review.</i>			
DOMAIN 3: Outcome			
A. Risk of Bias			
<i>Describe the outcome, how it was defined and determined, and the time interval between predictor assessment and outcome determination: Primary outcome measures were mortality and the combined end point of HF readmission or mortality within the follow-up period after the outpatient visit 1 month after discharge. Secondary end points encompassed the primary end points reached at 90, 180, and 365 days of follow-up.</i>			
	Dev	Val	
3.1 Was the outcome determined appropriately?	Y		
3.2 Was a pre-specified or standard outcome definition used?	Y		
3.3 Were predictors excluded from the outcome definition?	Y		
3.4 Was the outcome defined and determined in a similar way for all participants?	Y		
3.5 Was the outcome determined without knowledge of predictor information?	Y		
3.6 Was the time interval between predictor assessment and outcome determination appropriate?	Y		
Risk of bias introduced by the outcome or its determination	RISK:	Low	

	(low/ unclear)	high/	
<i>Rationale of bias rating: It has an adequate outcome definition that matches the research question in our review.</i>			
B. Applicability			
<i>At what time point was the outcome determined: It was determined during the follow-up in the future, during the time of hospitalization.</i>			
<i>If a composite outcome was used, describe the relative frequency/distribution of each contributing outcome:</i>			
Concern that the outcome, its definition, timing or determination do not match the review question	CONCERN: (low/ unclear)	high/	Low
<i>Rationale of applicability rating: The time in which the outcome is determined is consistent with the proposed research question.</i>			
DOMAIN 4: Analysis			
Risk of Bias			
<i>Describe numbers of participants, number of candidate predictors (for DEV only), outcome events and events per candidate predictor (for DEV only): In 309 out of 345 patients included in the PRIMA study, NT-proBNP levels at admission, discharge, and 1 month after hospital discharge were available. Patient characteristics at baseline and 1 month after hospital discharge are presented in the article. Candidate variables are not described</i>			
<i>Describe how the model was developed (predictor selection, optimism, risk groups, model performance): Data are presented as frequencies, mean 6 SD, or median (interquartile range [IQR]). Comparisons between groups were performed with the use of Fischer exact test for categoric data and 1-way analysis of variance or Kruskal-Wallis H test for continuous data, as appropriate. Glomerular filtration rate was estimated (eGFR) with the use of the Modified Diet in Renal Disease equation. Univariate Cox proportional hazard regression analysis was performed to assess clinical covariates associated with mortality. Spearman rank correlations where used to test correlations among the various NT-proBNP parameters. Multivariate Cox proportional hazard regression analysis was performed with the use of all covariates associated with outcome, except renal function and NT-proBNP concentration, to assess the clinical model. Variables were added in a stepwise fashion with $P < .05$ and $P < .1$ as the cutoffs for entry or retention, respectively. After assessment of the clinical model, renal function (eGFR <30 mL/min, eGFR 30-60 mL/min, or eGFR >60 mL/min) was added to form the reference model. To assess the independent prognostic value of NT-proBNP concentration on admission, at discharge, inpatient change, early outpatient change, and NT-proBNP concentration at 1 month after hospital discharge, these parameters were added to the reference model in a stepwise fashion to form the final NT-proBNP model. Model accuracy and discrimination were evaluated for both mortality and the combined end point of HF readmission or mortality within 1 year of follow-up by (i) c-statistic, a measure of the area under the receiver operating characteristic curve (AUC) and (ii) integrated discrimination improvement (IDI). Calculations were done with the use of IBM SPSS Statistics 21.0 (IBM, Armonk, New York) and Medcalc 13.3.3.0 (Medcalc Software, Ostend, Belgium).</i>			
<i>Describe whether and how the model was validated, either internally (e.g. bootstrapping, cross validation, random split sample) or externally (e.g. temporal validation, geographical validation, different setting, different type of participants): they don't mention the type of procedure for internal validation.</i>			
<i>Describe the performance measures of the model, e.g. (re)calibration, discrimination, (re)classification, net benefit: Model accuracy and discrimination for HF readmission and mortality reached a c-statistic of 0.85 (95% CI 0.81e0.90), with excellent calibration (Hosmer-Lemeshow: $P = .77$), and had significantly better model performance than models with 1 NT-proBNP parameter (IDI ranging from 6% to 13% [$P < .001$], improvement in c-statistic ranging from 0.3 to 0.7 [$P < .05$]). For mortality, the same trend was seen. Small changes in NT-proBNP concentration are associated with outcome. In multivariate analysis including the reference model, inpatient change in NT-proBNP, and NT-proBNP concentration at 1 month after discharge, early outpatient increase in NT-proBNP concentration $<30\%$ was associated with worse outcome compared with early</i>			

<p>outpatient decrease in NT-proBNP <30% (HR for mortality 2.05, 95% CI 1.02-4.13, Wald 4.1 [P = .04], HR for the combined end point 2.59, 95% CI 1.45-4.64, Wald 10.2 [P = .001]). Interestingly, there was no significant difference in mortality or the combined end point between patients with an early outpatient decrease of <30% vs ≥30% (HR for the combined end point 1.04, 95% CI 0.50-2.18, Wald 0.01; P = .914). Likewise, an increase in NTproBNP concentration of <30% yielded a clinically similar hazard for events compared with an increase ≥30% (HR for the combined end point 0.96, 95% CI 0.62-1.47, Wald 0.04; P = .837)</p>			
<p>Describe any participants who were excluded from the analysis: The patients who were excluded are not described, since the data obtained was from a selection from a database.</p>			
<p>Describe missing data on predictors and outcomes as well as methods used for missing data: Doesn't describe what was done with the lost data</p>			
		Dev	Val
4.1 Were there a reasonable number of participants with the outcome?		PN	
4.2 Were continuous and categorical predictors handled appropriately?		PY	
4.3 Were all enrolled participants included in the analysis?		N	
4.4 Were participants with missing data handled appropriately?		NI	
4.5 Was selection of predictors based on univariable analysis avoided?		Y	
4.6 Were complexities in the data (e.g. censoring, competing risks, sampling of controls) accounted for appropriately?		Y	
4.7 Were relevant model performance measures evaluated appropriately?		PY	
4.8 Were model overfitting and optimism in model performance accounted for?		NI	
4.9 Do predictors and their assigned weights in the final model correspond to the results from the reported multivariable analysis?		Y	
Risk of bias introduced by the analysis	RISK: (low/ high/ unclear)	High	
<p>Rationale of bias rating: There are missing data on the candidate variables and the way in which they were chosen, in addition to not mentioning the handling of the missing data.</p>			

Step 4: Overall assessment

<p>Use the following tables to reach overall judgments about risk of bias and concerns for applicability of the prediction model evaluation (development and/or validation) across all assessed domains.</p>	
<p>Reaching an overall judgment about risk of bias of the prediction model evaluation</p>	
Low risk of bias	<p>If all domains were rated low risk of bias.</p> <p>If a prediction model was developed without any external validation, and it was rated as <u>low risk of bias for all domains</u>, consider downgrading to high risk of bias. Such a model can only be considered as low risk of bias, if the development was based on a very large data set and included some form of internal validation.</p>
High risk of bias	<p>If at least one domain is judged to be at high risk of bias.</p>
Unclear risk of bias	<p>If an unclear risk of bias was noted in at least one domain and it was low risk for all other domains.</p>
<p>Complete for each evaluation of a distinct model.</p>	
<p>Reaching an overall judgment about applicability of the prediction model evaluation</p>	
Low concerns for applicability	<p>If low concerns for applicability for all domains, the prediction model evaluation is judged to have low concerns for applicability.</p>

High concerns for applicability	If high concerns for applicability for at least one domain, the prediction model evaluation is judged to have high concerns for applicability .
Unclear concerns for applicability	If unclear concerns (but no “high concern”) for applicability for at least one domain, the prediction model evaluation is judged to have unclear concerns for applicability overall.
Overall judgment about risk of bias and applicability of the prediction model evaluation	
Overall judgment of risk of bias	RISK: (low/ high/ unclear)
Summary of sources of potential bias: High	
Overall judgment of applicability	CONCERN: (low/ high/ unclear)
Summary of applicability concerns: Low	

PROBAST = Prediction model Risk Of Bias ASsessment Tool.

12. Formiga F, et al:

Step 2: Classify the type of prediction model evaluation

Classify the evaluation based on its aim			
Type of prediction study	PROBAST boxes to complete	Tick as appropriate	Definitions for type of prediction model study
Development only	Dev		Prediction model development without external validation. These studies may include internal validation methods such as bootstrapping and cross-validation techniques
Development and validation	Dev and Val		Prediction model development combined with external validation in other participants in the same article
Validation only	Val	X	External validation of existing (previously developed) model in other participants
<i>This table should be completed once for each publication being assessed and for each relevant outcome in your review.</i>			
Publication reference	Formiga F. Masip J. Chivite D. Corbella X. Applicability of the heart failure Readmission Risk score: A first European study. International Journal of Cardiology 236 (2017) 304–309. http://dx.doi.org/10.1016/j.ijcard.2017.02.024		
Models of interest	Readmission Risk score		

Outcome of interest	The applicability of the RR score outside the USA, for predicting not only future 30-day but also 90-day readmission in all patients aged 50 or more, who had been firstly admitted due to acute HF at a public tertiary care teaching hospital, in Barcelona, Spain.
----------------------------	---

Step 3: Assess risk of bias and applicability

DOMAIN 1: Participants			
A. Risk of Bias			
Describe the sources of data and criteria for participant selection: We retrieved administrative data regarding all admissions to our hospital within an 18-month-period (July 2013–December 2014) with HF as primary discharge diagnosis. Following this first selection, a thorough electronic medical record review was done to confirm that included patients truly fulfilled clinical criteria for acute HF, and that the index admission was the first ever hospitalization due to a first episode of acute HF.			
	De v		Val
1.1 Were appropriate data sources used, e.g. cohort, RCT or nested case-control study data?			Y
1.2 Were all inclusions and exclusions of participants appropriate?			Y
Risk of bias introduced by selection of participants	RISK: (low/ high/ unclear)		Low
<i>Rationale of bias rating:</i>			
Inclusion criteria was correctly mentioned.			
B. Applicability			
Describe included participants, setting and dates: A total of 679 patients met the inclusion criteria and were included in the study. The mean age was 78.1 ± 9 years (range 52–102), and 377 (55.5%) of them were females. With respect to the Components of Readmission risk Score: 258 (38.1%) patients had a prior history of diabetes, 164 (24.2%) CAD, 101 (14.9%) previous PCI, 61 (9%) aortic stenosis, 102 (15%) stroke, 138 (20.3%) COPD and 52 (7.7%) dementia. The average values of patients' vital signs and laboratory data presenting on admission were: systolic blood pressure 139±54mmHg; heart rate 86.2±31/min and respiratory rate 20,8 ± 6/min, plasma sodium 188.8 ± 4 mEq/L, glucose 138.72 ± 66 mg/dL, hematocrit (HCT) 36.2 ± 5.1, creatinine 1.32 ± 1.1 mg/L, and blood urea nitrogen (BUN) 47.7 ± 21 mg/L.			
Concern that the included participants and setting do not match the review question	CONCERN: (low/ high/ unclear)		Low
<i>Rationale of applicability rating: Information from patients were correctly mentioned.</i>			

DOMAIN 2: Predictors			
A. Risk of Bias			
List and describe predictors included in the final model, e.g. definition and timing of assessment: age, sex, history of diabetes, previous HF (in our case we considered no for all patients), coronary artery disease (CAD), previous percutaneous coronary intervention (PCI), aortic stenosis, stroke, chronic obstructive pulmonary disease (COPD), and dementia prior diagnosis. Vital signs included on admission were systolic blood pressure, heart rate, and respiratory rate. Laboratory data included on admission were plasma sodium, blood urea nitrogen (BUN), hematocrit (HCT), creatinine, and glucose; left ventricular ejection fraction (LVEF) as determined by means of echocardiography was also included.			
	De v		Val
2.1 Were predictors defined and assessed in a similar way for all participants?			Y

2.2 Were predictor assessments made without knowledge of outcome data?			Y
2.3 Are all predictors available at the time the model is intended to be used?			Y
Risk of bias introduced by predictors or their assessment	RISK: (low/ high/ unclear)		Low
<i>Rationale of bias rating: The article has a good selection of variables combining variables known as risk factors in AHF.</i>			
B. Applicability			
Concern that the definition, assessment or timing of predictors in the model do not match the review question	CONCERN: (low/ high/ unclear)		Low
<i>Rationale of applicability rating: They use a simple group of variables, easy to get for develop one usefull and very applicable model.</i>			
DOMAIN 3: Outcome			
A. Risk of Bias			
<i>Describe the outcome, how it was defined and determined, and the time interval between predictor assessment and outcome determination: The applicability of the RR score outside the USA, for predicting not only future 30-day but also 90-day readmission in all patients aged 50 or more, who had been firstly admitted due to acute HF at a public tertiary care teaching hospital, in Barcelona, Spain. Readmission was defined as any unplanned hospital readmission for any cause within the first 30 days after discharge due to a first acute HF decompensation — primary outcome, and within the first 90 days after discharge — secondary outcome.</i>			
		Dev	Val
3.1 Was the outcome determined appropriately?			Y
3.2 Was a pre-specified or standard outcome definition used?			Y
3.3 Were predictors excluded from the outcome definition?			Y
3.4 Was the outcome defined and determined in a similar way for all participants?			Y
3.5 Was the outcome determined without knowledge of predictor information?			NI
3.6 Was the time interval between predictor assessment and outcome determination appropriate?			Y
Risk of bias introduced by the outcome or its determination	RISK: (low/ high/ unclear)		Low
<i>Rationale of bias rating: It was very well defined outcome, with good model</i>			
B. Applicability			
<i>At what time point was the outcome determined: at one time in the future</i>			
<i>If a composite outcome was used, describe the relative frequency/distribution of each contributing outcome:</i>			
Concern that the outcome, its definition, timing or determination do not match the review question	CONCERN: (low/ high/ unclear)		Low
<i>Rationale of applicability rating: It was well defined the outcome and it match with the review question.</i>			
DOMAIN 4: Analysis			
Risk of Bias			
<i>Describe numbers of participants, number of candidate predictors (for DEV only), outcome events and events per candidate predictor (for DEV only): A total of 679 patients met the inclusion criteria and were included in the</i>			

<p>study. The mean age was 78.1 ± 9 years (range 52–102), and 377 (55.5%) of them were females. There were no candidate predictors or EPV.</p>		
<p>Describe how the model was developed (predictor selection, optimism, risk groups, model performance): The average RR score for these patients was significantly higher (22.7 vs. 20.1; $p < 0.0001$) than that of those who did not experience an early readmission. There were no differences regarding age and gender between these two groups of patients. the multivariate logistic regression results associated with 30- day readmission with each component of the RR score evaluated; sodium value was the only significantly associated with that outcome (odds ratio 0.882). In logistic regression with the use of 30-day all-cause read- mission as an outcome and the RR score as an independent variable, for each 1-point increase in the RR score, the odds of 30-day readmission increased by 9.3%. After 90 days of discharge, 98 patients (14.4%) had experienced a new readmission for any cause. The average RR score of these patients was again significantly higher than that of patients who were not readmitted (21.9 vs. 20.1; $p < 0.001$). Both groups were also similar regarding age and sex. the multivariate logistic regression results for 90-day hospital admissions considering each component of the RR score evaluated. Creatinine value was the only variable reaching statistical significance (odds ratio 1.006). In logistic regression, for each 1-point increase in the RR score, the odds of 90-day readmission increased by 7.2%. Results also showed a significant increase in 90-day readmission percentages according to increasing RR score quartiles. For patients in the highest quartile, the risk of 90-day readmission was much higher than that for patients in the lowest quartile (41.8% vs. 16.3%). The ROC curve of the RR score for all-cause 90-day read- mission. The area under the curve was the 0.621 (95% CI 0.560–0.681). There were no significant differences in ROC curves according to whether LVEF was assessed or not during admission in in patients readmitted or not at 90 days (0.609 vs. 0.602).</p>		
<p>Describe whether and how the model was validated, either internally (e.g. bootstrapping, cross validation, random split sample) or externally (e.g. temporal validation, geographical validation, different setting, different type of participants): It was validated externally</p>		
<p>Describe the performance measures of the model, e.g. (re)calibration, discrimination, (re)classification, net benefit: In our independent applicability study, the first in Europe and outside from the USA and Canada, in a HF patient population and health care environment with notably differences from the originally tested by YNHSC/CORE, the C-statistic value was of 0.64 for predicting unplanned admissions within 30 days after discharge for a first HF admission, and 0.62 within 90 days. Therefore, although modest at best, our results confirm the idea that the RR score C-statistic is better than random chance predicting all-cause of unplanned readmissions, not only within 30 days but also 90 days after a first HF hospitalization.</p>		
<p>Describe any participants who were excluded from the analysis: all those who had been discharged with a primary or secondary diagnosis of HF prior to the index admission. Patients younger than 50 years of age, those who had been discharged directly home within 24 h or transferred to other acute care hospitals from the Emergency Department, those in stage V chronic kidney disease (CKD) undergoing kidney replacement therapies, those who had received a kidney, liver or heart transplant, those whose acute HF was secondary to an acute coronary syndrome and those who had been subject to prior blood transfusions with palliative intention were also excluded from the study</p>		
<p>Describe missing data on predictors and outcomes as well as methods used for missing data: There were no missing data except for echocardiography data in some patients.</p>		
	Dev	Val
4.1 Were there a reasonable number of participants with the outcome?		Y
4.2 Were continuous and categorical predictors handled appropriately?		Y
4.3 Were all enrolled participants included in the analysis?		Y
4.4 Were participants with missing data handled appropriately?		Y
4.5 Was selection of predictors based on univariable analysis avoided?		Y
4.6 Were complexities in the data (e.g. censoring, competing risks, sampling of controls) accounted for appropriately?		NI

4.7 Were relevant model performance measures evaluated appropriately?		Y
4.8 Were model overfitting and optimism in model performance accounted for?		Y
4.9 Do predictors and their assigned weights in the final model correspond to the results from the reported multivariable analysis?		Y
Risk of bias introduced by the analysis	RISK: (low/ high/ unclear)	Low
<i>Rationale of bias rating: Describe how they get the score</i>		

Step 4: Overall assessment

Use the following tables to reach overall judgments about risk of bias and concerns for applicability of the prediction model evaluation (development and/or validation) across all assessed domains.

Reaching an overall judgment about risk of bias of the prediction model evaluation	
Low risk of bias	If all domains were rated low risk of bias. If a <u>prediction model was developed without any external validation</u> , and it was rated as <u>low risk of bias for all domains</u> , consider downgrading to high risk of bias . Such a model can only be considered as low risk of bias, if the development was based on a very large data set <u>and</u> included some form of internal validation.
High risk of bias	If at least one domain is judged to be at high risk of bias .
Unclear risk of bias	If an unclear risk of bias was noted in at least one domain and it was low risk for all other domains.

Complete for each evaluation of a distinct model.

Reaching an overall judgment about applicability of the prediction model evaluation	
Low concerns for applicability	If low concerns for applicability for all domains, the prediction model evaluation is judged to have low concerns for applicability .
High concerns for applicability	If high concerns for applicability for at least one domain, the prediction model evaluation is judged to have high concerns for applicability .
Unclear concerns for applicability	If unclear concerns (but no “high concern”) for applicability for at least one domain, the prediction model evaluation is judged to have unclear concerns for applicability overall.

Overall judgment about risk of bias and applicability of the prediction model evaluation		
Overall judgment of risk of bias	RISK: (low/ high/ unclear)	Low
<i>Summary of sources of potential bias: Low</i>		
Overall judgment of applicability	CONCERN: (low/ high/ unclear)	Low

Summary of applicability concerns: Low

PROBAST = Prediction model Risk Of Bias ASsessment Tool.

13. Garcia-Gutierrez JS, et al:

Step 2: Classify the type of prediction model evaluation

Classify the evaluation based on its aim			
Type of prediction study	PROBAST boxes to complete	Tick as appropriate	Definitions for type of prediction model study
Development only	Dev		Prediction model development without external validation. These studies may include internal validation methods such as bootstrapping and cross-validation techniques
Development and validation	Dev and Val	X	Prediction model development combined with external validation in other participants in the same article
Validation only	Val		External validation of existing (previously developed) model in other participants
<i>This table should be completed once for each publication being assessed and for each relevant outcome in your review.</i>			
Publication reference	Garcia-Gutierrez JS, Quintana M, Antón-Ladislao A, Gallardo M. Pulido E. Creation and validation of the acute heart failure risk score: AHFRS. Intern Emerg Med (2017) 12:1197–1206 DOI 10.1007/s11739-016-1541-4		
Models of interest	AHFRS		
Outcome of interest	Create and validate a scale to assess severity in patients with AHF who visit the ED, regardless of whether they are admitted or discharged, in terms of poor course (death and complications) during admission or up to seven days after the index episode. We will compare the predictive accuracy of this new model and the accuracy of other models validated in other regions and based on administrative data. We also explored the additive predictive value of high sensitivity (hs)-TNT and NT-proBNP in the final model.		

Step 3: Assess risk of bias and applicability

DOMAIN 1: Participants
A. Risk of Bias

Describe the sources of data and criteria for participant selection: This prospective cohort study recruited patients with symptoms of AHF who were seen in the ED of any of the three hospitals in the Basque National Health Service between April 2011 and April 2013. Patients were eligible for inclusion in the study if they were diagnosed with heart failure: Acute decompensated heart failure, hypertensive AHF, pulmonary oedema, cardiogenic shock, isolated right HF, and Acute Coronary Syndrome and Heart failure			
		De v	Val
1.1 Were appropriate data sources used, e.g. cohort, RCT or nested case-control study data?		Y	Y
1.2 Were all inclusions and exclusions of participants appropriate?		Y	Y
Risk of bias introduced by selection of participants	RISK: (low/ high/ unclear)	Low	Low
<i>Rationale of bias rating:</i>			
Inclusion criteria was correctly mentioned. No information clearly defined for exclusion criteria			
B. Applicability			
<i>Describe included participants, setting and dates: A total of 2032 patients were recruited for the study. Of these, we excluded 176 (8.66 %), mainly because there was no confirmation of ADHF (81 patients), though also due to the main diagnosis being some complex comorbidity rather than heart failure (34 patients), among other reasons. Finally, 1824 patients were included in the study, 1467 (79.4 %) of them were admitted to hospital and 389 (20.96 %) were discharged home.</i>			
Concern that the included participants and setting do not match the review question	CONCERN: (low/ high/ unclear)	Low	Low
<i>Rationale of applicability rating: Information from patients were correctly mentioned.</i>			

DOMAIN 2: Predictors			
A. Risk of Bias			
<i>List and describe predictors included in the final model, e.g. definition and timing of assessment: We collected data on variables related to the background of the patient (medical history). Oedema (yes vs. no), Visits to ED and/or admissions in last 2 years, Glucose level (≥ 122 vs. < 122 mg/dl), Blood urea nitrogen level (≥ 16 vs. < 16 mg/dl)</i>			
		De v	Val
2.1 Were predictors defined and assessed in a similar way for all participants?		Y	Y
2.2 Were predictor assessments made without knowledge of outcome data?		Y	Y
2.3 Are all predictors available at the time the model is intended to be used?		Y	Y
Risk of bias introduced by predictors or their assessment	RISK: (low/ high/ unclear)	Low	Low
<i>Rationale of bias rating: The article has a good selection of variables combining variables known as risk factors in AHF.</i>			
B. Applicability			

Concern that the definition, assessment or timing of predictors in the model do not match the review question	CONCERN: (low/ high/ unclear)	Low	Low
<i>Rationale of applicability rating: They use a simple group of variables, easy to get for develop one usefull and very applicable model.</i>			
DOMAIN 3: Outcome			
A. Risk of Bias			
Describe the outcome, how it was defined and determined, and the time interval between predictor assessment and outcome determination: Create and validate a scale to assess severity in patients with AHF who visit the ED, regardless of whether they are admitted or discharged, in terms of poor course (death and complications) during admission or up to seven days after the index episode. We will compare the predictive accuracy of this new model and the accuracy of other models validated in other regions and based on administrative data. We also explored the additive predictive value of high sensitivity (hs)-TNT and NT-proBNP in the final model. "Poor course during the acute phase" is a patient centered composite endpoint that included: (a) death in-hospital for admitted patients and during the first week following the ED visit for discharged patients), (b) admission to an intensive care unit; need for invasive mechanical ventilaion, or need for non-invasive mechanical ventilation dur- ing their stay in the emergency department.			
		Dev	Val
3.1 Was the outcome determined appropriately?		Y	Y
3.2 Was a pre-specified or standard outcome definition used?		Y	Y
3.3 Were predictors excluded from the outcome definition?		Y	Y
3.4 Was the outcome defined and determined in a similar way for all participants?		Y	Y
3.5 Was the outcome determined without knowledge of predictor information?		Y	NI
3.6 Was the time interval between predictor assessment and outcome determination appropriate?		Y	Y
Risk of bias introduced by the outcome or its determination	RISK: (low/ high/ unclear)	Low	Low
<i>Rationale of bias rating: It was very well defined outcome, with good model</i>			
B. Applicability			
<i>At what time point was the outcome determined: at one time in the future</i>			
<i>If a composite outcome was used, describe the relative frequency/distribution of each contributing outcome:</i>			
Concern that the outcome, its definition, timing or determination do not match the review question	CONCERN: (low/ high/ unclear)	Low	Low
<i>Rationale of applicability rating: It was well defined the outcome and it match with the review question.</i>			
DOMAIN 4: Analysis			
Risk of Bias			
Describe numbers of participants, number of candidate predictors (for DEV only), outcome events and events per candidate predictor (for DEV only): 1824 patients were included in the study, 1467 (79.4 %) of them were admitted to hospital and 389 (20.96 %) were discharged home. No information about candidate predictors or EPV.			

<p>Describe how the model was developed (predictor selection, optimism, risk groups, model performance): We developed the acute heart failure risk score (AHFRS) by assigning a weight to each risk factor category based on the b parameter from the multivariate logistic regression. Then, we added the weights of each of the risk factors presented by a patient, with a higher score corresponding to a higher likelihood of poor course. Considering the optimal classifier and taking into account the distribution of poor outcomes, four AHFRS severity categories were created for the model: very low, low, intermediate and high risk. We created a logistic regression model using our score (both as a continuous and as a categorical variable) as an independent variable, and calculated the AUCs. Final models were also adjusted for the treating hospital to assess whether that affected the results.</p>			
<p>Describe whether and how the model was validated, either internally (e.g. bootstrapping, cross validation, random split sample) or externally (e.g. temporal validation, geographical validation, different setting, different type of participants): They develop the score and validate it.</p>			
<p>Describe the performance measures of the model, e.g. (re)calibration, discrimination, (re)classification, net benefit: This model obtained AUC of 0.83 (CI 0.79-0.88). The risk score was significantly associated with the likelihood of poor course (OR 1.28; 95 % CI 1.204–1.359; P<0.0001), and was well calibrated. It showed good discrimination (AUC 0.84; 95 % CI 0.79–0.88).</p>			
<p>Describe any participants who were excluded from the analysis: They excluded 176 (8.66 %), mainly because there was no confirmation of ADHF (81 patients), though also due to the main diagnosis being some complex comorbidity rather than heart failure (34 patients), among other reasons.</p>			
<p>Describe missing data on predictors and outcomes as well as methods used for missing data: Do not present any missing data</p>			
		Dev	Val
4.1 Were there a reasonable number of participants with the outcome?		Y	Y
4.2 Were continuous and categorical predictors handled appropriately?		Y	Y
4.3 Were all enrolled participants included in the analysis?		Y	Y
4.4 Were participants with missing data handled appropriately?		NI	Y
4.5 Was selection of predictors based on univariable analysis avoided?		Y	Y
4.6 Were complexities in the data (e.g. censoring, competing risks, sampling of controls) accounted for appropriately?		NI	NI
4.7 Were relevant model performance measures evaluated appropriately?		Y	Y
4.8 Were model overfitting and optimism in model performance accounted for?		Y	Y
4.9 Do predictors and their assigned weights in the final model correspond to the results from the reported multivariable analysis?		Y	NI
Risk of bias introduced by the analysis	RISK: (low/ high/ unclear)	Low	Low
<p>Rationale of bias rating: Describe how they get the score</p>			

Step 4: Overall assessment

Use the following tables to reach overall judgments about risk of bias and concerns for applicability of the prediction model evaluation (development and/or validation) across all assessed domains.

Reaching an overall judgment about risk of bias of the prediction model evaluation	
Low risk of bias	If all domains were rated low risk of bias. If a prediction model was developed without any external validation, and it was rated as <u>low risk of bias for all domains</u> , consider downgrading to high risk of bias . Such a model can only be considered as low risk of bias, if the development was based on a very large data set and included some form of internal validation.
High risk of bias	If at least one domain is judged to be at high risk of bias .
Unclear risk of bias	If an unclear risk of bias was noted in at least one domain and it was low risk for all other domains.

Complete for each evaluation of a distinct model.

Reaching an overall judgment about applicability of the prediction model evaluation	
Low concerns for applicability	If low concerns for applicability for all domains, the prediction model evaluation is judged to have low concerns for applicability .
High concerns for applicability	If high concerns for applicability for at least one domain, the prediction model evaluation is judged to have high concerns for applicability .
Unclear concerns for applicability	If unclear concerns (but no “high concern”) for applicability for at least one domain, the prediction model evaluation is judged to have unclear concerns for applicability overall.

Overall judgment about risk of bias and applicability of the prediction model evaluation		
Overall judgment of risk of bias	RISK: (low/ high/ unclear)	Low
<i>Summary of sources of potential bias: Low</i>		
Overall judgment of applicability	CONCERN: (low/ high/ unclear)	Low
<i>Summary of applicability concerns: Low</i>		

PROBAST = Prediction model Risk Of Bias ASsessment Tool.

14. Gil V, et al

Step 2: Classify the type of prediction model evaluation

Classify the evaluation based on its aim			
Type of prediction study	PROBAST boxes to complete	Tick as appropriate	Definitions for type of prediction model study
Development only	Dev		Prediction model development without external validation. These studies may include internal validation methods such as bootstrapping and cross-validation techniques
Development and validation	Dev and Val		Prediction model development combined with external validation in other participants in the same article
Validation only	Val	X	External validation of existing (previously developed) model in other participants
<i>This table should be completed once for each publication being assessed and for each relevant outcome in your review.</i>			
Publication reference	Gil V. Miro O. Schull M, Llorens P, Herrero-Puente P. Emergency Heart Failure Mortality Risk Grade score performance for 7-day mortality prediction in patients with heart failure attended at the emergency department: validation in a Spanish cohort. European Journal of Emergency Medicine 2018, Vol 25 No 3 DOI: 10.1097/MEJ.0000000000000422		
Models of interest	EHMRG scale		
Outcome of interest	To measure the performance of the EHMRG scale in a broader spectrum of patients diagnosed with ADHF in Spanish EDs, including palliative and non- palliative patients, to externally validate this scale.		

Step 3: Assess risk of bias and applicability

DOMAIN 1: Participants
A. Risk of Bias
Describe the sources of data and criteria for participant selection: The EAHFE registry is formed by consecutive AHF patients diagnosed at 29 Spanish EDs in both university and community hospitals representing all the areas of the country throughout Spain. Only patients included between 1 November and

31 December 2011 (EAHFE-3 spot) were considered because it was the only spot in which patient arrival to the ED. The diagnosis of ADHF was made using the Framingham diagnostic criteria because, despite having some limitations, these are the criteria most commonly used on clinical grounds			
		De v	Val
1.1 Were appropriate data sources used, e.g. cohort, RCT or nested case-control study data?			Y
1.2 Were all inclusions and exclusions of participants appropriate?			NI
Risk of bias introduced by selection of participants	RISK: (low/ high/ unclear)		Uncle ar
<i>Rationale of bias rating:</i>			
Inclusion or exclusion criteria were not well defined.			
B. Applicability			
Describe included participants, setting and dates: The study population consisted of patients with AHF. Only patients included between 1 November and 31 December 2011. Of the 3414 patients included in the EAHFE-3 cohort, 1553 were finally included in the present study and constituted the EAHFE group. The main reason for patient exclusion was the lack of troponin determination in the ED, having been performed in only 2031 cases (59.5%).			
Concern that the included participants and setting do not match the review question	CONCERN: (low/ high/ unclear)		Low
<i>Rationale of applicability rating: Information from patients were correctly mentioned.</i>			

DOMAIN 2: Predictors			
A. Risk of Bias			
<i>List and describe predictors included in the final model, e.g. definition and timing of assessment: They use the EHMRG score, previously developed. Age, transport by EMS, SBP, Heart Rate, oxygen saturation, creatinine, potassium, elevated troponin level, Active cancer and use of Metazolone at home.</i>			
		De v	Val
2.1 Were predictors defined and assessed in a similar way for all participants?			Y
2.2 Were predictor assessments made without knowledge of outcome data?			Y
2.3 Are all predictors available at the time the model is intended to be used?			Y
Risk of bias introduced by predictors or their assessment	RISK: (low/ high/ unclear)		Low
<i>Rationale of bias rating: The article has a good selection of variables combining variables known as risk factors in AHF.</i>			
B. Applicability			

Concern that the definition, assessment or timing of predictors in the model do not match the review question	CONCERN: (low/ high/ unclear)		Low
<i>Rationale of applicability rating: They use a simple group of variables, easy to get for develop one usefull and very applicable model.</i>			
DOMAIN 3: Outcome			
A. Risk of Bias			
Describe the outcome, how it was defined and determined, and the time interval between predictor assessment and outcome determination: To measure the performance of the EHMGR scale in a broader spectrum of patients diagnosed with ADHF in Spanish EDs, including palliative and non-palliative patients, to externally validate this scale.			
		Dev	Val
3.1 Was the outcome determined appropriately?			Y
3.2 Was a pre-specified or standard outcome definition used?			Y
3.3 Were predictors excluded from the outcome definition?			Y
3.4 Was the outcome defined and determined in a similar way for all participants?			Y
3.5 Was the outcome determined without knowledge of predictor information?			NI
3.6 Was the time interval between predictor assessment and outcome determination appropriate?			Y
Risk of bias introduced by the outcome or its determination	RISK: (low/ high/ unclear)		Low
<i>Rationale of bias rating: It was very well defined outcome, with good model</i>			
B. Applicability			
<i>At what time point was the outcome determined: at one time in the future</i>			
<i>If a composite outcome was used, describe the relative frequency/distribution of each contributing outcome:</i>			
Concern that the outcome, its definition, timing or determination do not match the review question	CONCERN: (low/ high/ unclear)		Low
<i>Rationale of applicability rating: It was well defined the outcome and it match with the review question.</i>			

DOMAIN 4: Analysis			
Risk of Bias			
Describe numbers of participants, number of candidate predictors (for DEV only), outcome events and events per candidate predictor (for DEV only): 1553 were finally included in the present study and constituted the EAHFE group. No candidate predictors or EPV.			
Describe how the model was developed (predictor selection, optimism, risk groups, model performance): They used the EHMRG score previously developed, with good performance.			
Describe whether and how the model was validated, either internally (e.g. bootstrapping, cross validation, random split sample) or externally (e.g. temporal validation, geographical validation, different setting, different type of participants): They externally validated the score in this population.			
Describe the performance measures of the model, e.g. (re)calibration, discrimination, (re)classification, net benefit: The mean EHMRG score for the EAHFE group was 54.6 points (range: -123.6 to 261.9 points) and fit a normal distribution pattern ($P = 0.27$; Fig. 2a). The 7-day mortality rate following ED consultation was 5.5% (85 deaths). The AUC ROC of the EHMRG scale was 0.741 for the prediction of 7-day mortality in the EAHFE group of ADHF patients diagnosed at Spanish EDs.			
On carrying out sensitivity analyses of the discriminative capacity of the EHMRG scale in the EAHFE population, we did not find statistically significant differences with respect to the original discriminative capacity found by us.			
Describe any participants who were excluded from the analysis: The main reason for patient exclusion was the lack of troponin determination in the ED			
Describe missing data on predictors and outcomes as well as methods used for missing data: imputing missing values for troponin results as a 'normal' value			
		Dev	Val
4.1 Were there a reasonable number of participants with the outcome?			Y
4.2 Were continuous and categorical predictors handled appropriately?			Y
4.3 Were all enrolled participants included in the analysis?			Y
4.4 Were participants with missing data handled appropriately?			N
4.5 Was selection of predictors based on univariable analysis avoided?			Y
4.6 Were complexities in the data (e.g. censoring, competing risks, sampling of controls) accounted for appropriately?			NI
4.7 Were relevant model performance measures evaluated appropriately?			Y
4.8 Were model overfitting and optimism in model performance accounted for?			Y
4.9 Do predictors and their assigned weights in the final model correspond to the results from the reported multivariable analysis?			NI
Risk of bias introduced by the analysis		RISK: (low/ high/ unclear)	High
Rationale of bias rating: Describe how they get the score			

Step 4: Overall assessment

Use the following tables to reach overall judgments about risk of bias and concerns for applicability of the prediction model evaluation (development and/or validation) across all assessed domains.

Reaching an overall judgment about risk of bias of the prediction model evaluation	
Low risk of bias	If all domains were rated low risk of bias. If a prediction model was developed without any external validation, and it was rated as <u>low risk of bias for all domains</u> , consider downgrading to high risk of bias . Such a model can only be considered as low risk of bias, if the development was based on a very large data set and included some form of internal validation.
High risk of bias	If at least one domain is judged to be at high risk of bias .
Unclear risk of bias	If an unclear risk of bias was noted in at least one domain and it was low risk for all other domains.

Complete for each evaluation of a distinct model.

Reaching an overall judgment about applicability of the prediction model evaluation	
Low concerns for applicability	If low concerns for applicability for all domains, the prediction model evaluation is judged to have low concerns for applicability .
High concerns for applicability	If high concerns for applicability for at least one domain, the prediction model evaluation is judged to have high concerns for applicability .
Unclear concerns for applicability	If unclear concerns (but no “high concern”) for applicability for at least one domain, the prediction model evaluation is judged to have unclear concerns for applicability overall.

Overall judgment about risk of bias and applicability of the prediction model evaluation		
Overall judgment of risk of bias	RISK: (low/ high/ unclear)	High
Summary of sources of potential bias: High		
Overall judgment of applicability	CONCERN: (low/ high/ unclear)	Low
Summary of applicability concerns: Low		

PROBAST = Prediction model Risk Of Bias ASsessment Tool.

15. Hsieh M, et al

Step 2: Classify the type of prediction model evaluation

Classify the evaluation based on its aim

Type of prediction study	PROBAST boxes to complete	Tick as appropriate	Definitions for type of prediction model study
Development only	Dev		Prediction model development without external validation. These studies may include internal validation methods such as bootstrapping and cross-validation techniques
Development and validation	Dev and Val	X	Prediction model development combined with external validation in other participants in the same article
Validation only	Val		External validation of existing (previously developed) model in other participants
<i>This table should be completed once for each publication being assessed and for each relevant outcome in your review.</i>			
Publication reference	Hsieh M. uble T. Donald M. Yealy D. Validation of the Acute Heart Failure Index. Department of Emergency Medicine, University of Pittsburgh, Pittsburgh, PA. Volume 51 No. 1: January 2008. Annals of Emergency Medicine. doi:10.1016/j.annemergmed.2007.07.026		
Models of interest	Acute Heart Failure Index		
Outcome of interest	Validate a clinical prediction rule prognostic of short-term fatal and inpatient nonfatal outcomes for heart failure patients admitted through the emergency department.		

Step 3: Assess risk of bias and applicability

DOMAIN 1: Participants			
A. Risk of Bias			
Describe the sources of data and criteria for participant selection: Our study cohort consisted of a random computer-generated sample of 5,000 patients discharged from all Pennsylvania general acute care hospitals for each of the calendar years 2003 and 2004 with a diagnosis of heart failure, defined as an <i>International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM)</i> ¹¹ hospital primary discharge diagnosis code consistent with heart failure (398.91, 402.01, 402.11, 402.91, 404.01, 404.03, 404.11, 404.13, 404.91, 404.93, 425.4, 428, 428.0, 428.1, 428.2x-428.4x, or 428.9). We included patients with these diagnoses if they were 18 years of age or older, Pennsylvania residents, and hospitalized from the ED during the study period.			
		Dev	Val
1.1 Were appropriate data sources used, e.g. cohort, RCT or nested case-control study data?			Y
1.2 Were all inclusions and exclusions of participants appropriate?			Y
Risk of bias introduced by selection of participants		RISK: (low/unclear) high/	Low
<i>Rationale of bias rating:</i>			
Inclusion and exclusion criteria were correctly defined.			
B. Applicability			

Describe included participants, setting and dates: From the initial group of 10,000 patients, 241 were excluded because their Pennsylvania Health Care Cost Containment Council records could not be linked to their MediQual-Atlas records, 364 had no documented ED vital signs, 4 had no key clinical findings from the day of admission onward, and 376 were transferred to another short-term general hospital for inpatient care. An additional 631 records were excluded because those patients were included in the derivation cohort, or the 2003 cohort in the case of 2004 patients. The final sample size for the validation group was 8,384 patients. Eighty-one percent of the patients were 65 years or older. The most common comorbid existing illnesses of these patients were a history of heart failure (68.6%), diabetes (38.3%), or lung disease (33.6%). Radiographic evidence of pulmonary congestion was present in 50.8% of patients. Electrocardiographic abnormalities documented at the time of presentation included acute myocardial infarction (17.2%) or myocardial ischemia not known to be old (13.2%). Overall, 318 (3.8%; 95% CI 3.4% to 4.2%) patients died during hospitalization, and 490 (5.8%; 95% CI 5.3% to 6.3%) survived to hospital discharge after experiencing a serious medical complication. With respect to the secondary study outcome, 736 (8.8%; 95% CI 8.2% to 9.4%) patients died within 30 days of the index hospitalization.

Concern that the included participants and setting do not match the review question

CONCERN:
(low/ high/
unclear)

Low

Rationale of applicability rating: Information from patients were correctly mentioned.

DOMAIN 2: Predictors

A. Risk of Bias

List and describe predictors included in the final model, e.g. definition and timing of assessment:

		De v	Val
2.1 Were predictors defined and assessed in a similar way for all participants?			Y
2.2 Were predictor assessments made without knowledge of outcome data?			Y
2.3 Are all predictors available at the time the model is intended to be used?			Y
Risk of bias introduced by predictors or their assessment	RISK: (low/ high/ unclear)		Low
<i>Rationale of bias rating: The article has a good selection of variables combining variables known as risk factors in AHF.</i>			
B. Applicability			
Concern that the definition, assessment or timing of predictors in the model do not match the review question	CONCERN: (low/ high/ unclear)		Low
<i>Rationale of applicability rating: They use a simple group of variables, easy to get for develop one usefull and very applicable model.</i>			
DOMAIN 3: Outcome			
A. Risk of Bias			
Describe the outcome, how it was defined and determined, and the time interval between predictor assessment and outcome determination: Validate a clinical prediction rule prognostic of short-term fatal and inpatient nonfatal outcomes for heart failure patients admitted through the emergency department. The primary outcome was death or a serious medical complication before hospital discharge. We described performance of the prediction rule for inpatient death alone and 30-day mortality.			

		Dev	Val
3.1 Was the outcome determined appropriately?			Y
3.2 Was a pre-specified or standard outcome definition used?			Y
3.3 Were predictors excluded from the outcome definition?			Y
3.4 Was the outcome defined and determined in a similar way for all participants?			Y
3.5 Was the outcome determined without knowledge of predictor information?			NI
3.6 Was the time interval between predictor assessment and outcome determination appropriate?			Y
Risk of bias introduced by the outcome or its determination	RISK: (low/ high/ unclear)		Low
<i>Rationale of bias rating: It was very well defined outcome, with good model</i>			
B. Applicability			
<i>At what time point was the outcome determined: at one time in the future</i>			
<i>If a composite outcome was used, describe the relative frequency/distribution of each contributing outcome:</i>			
Concern that the outcome, its definition, timing or determination do not match the review question	CONCERN: (low/ high/ unclear)		Low
<i>Rationale of applicability rating: It was well defined the outcome and it match with the review question.</i>			
DOMAIN 4: Analysis			
Risk of Bias			
<i>Describe numbers of participants, number of candidate predictors (for DEV only), outcome events and events per candidate predictor (for DEV only): The final sample size for the validation group was 8,384 patients. No information of candidate predictors or EPV was defined.</i>			
<i>Describe how the model was developed (predictor selection, optimism, risk groups, model performance): They use the AHF index score, which was previously presented.</i>			
<i>Describe whether and how the model was validated, either internally (e.g. bootstrapping, cross validation, random split sample) or externally (e.g. temporal validation, geographical validation, different setting, different type of participants): no mentioned.</i>			
<i>Describe the performance measures of the model, e.g. (re)calibration, discrimination, (re)classification, net benefit: They validated externally in this population from Pennsylvania Health Care Cost Containment Council records.</i>			
<i>Describe any participants who were excluded from the analysis: excluded patients without ED vital signs such as pulse, systolic blood pressure, and respiratory rate to ensure that patients were evaluated in the ED, as well as patients without any key clinical findings from the day of admission onward to ensure that patients were truly admitted. Patients transferred to another short-term general hospital for inpatient care were also excluded because we could not evaluate their entire hospitalization. From the initial group of 10,000 patients, 241 were excluded because their Pennsylvania Health Care Cost Containment Council records could not be linked to their MediQual-Atlas records, 364 had no documented ED vital signs, 4 had no key clinical findings from the day of admission onward, and 376 were transferred to another short-term general hospital for inpatient care.</i>			

An additional 631 records were excluded because those patients were included in the derivation cohort, or the 2003 cohort in the case of 2004 patients			
<i>Describe missing data on predictors and outcomes as well as methods used for missing data: They do not explain very well how they handled the missing data.</i>			
		Dev	Val
4.1 Were there a reasonable number of participants with the outcome?			Y
4.2 Were continuous and categorical predictors handled appropriately?			Y
4.3 Were all enrolled participants included in the analysis?			Y
4.4 Were participants with missing data handled appropriately?			N
4.5 Was selection of predictors based on univariable analysis avoided?			Y
4.6 Were complexities in the data (e.g. censoring, competing risks, sampling of controls) accounted for appropriately?			NI
4.7 Were relevant model performance measures evaluated appropriately?			Y
4.8 Were model overfitting and optimism in model performance accounted for?			Y
4.9 Do predictors and their assigned weights in the final model correspond to the results from the reported multivariable analysis?			NI
Risk of bias introduced by the analysis	RISK: (low/ high/ unclear)		Unclear
<i>Rationale of bias rating: Describe how they get the score</i>			

Step 4: Overall assessment

Use the following tables to reach overall judgments about risk of bias and concerns for applicability of the prediction model evaluation (development and/or validation) across all assessed domains.

Reaching an overall judgment about risk of bias of the prediction model evaluation	
Low risk of bias	If all domains were rated low risk of bias. If a <u>prediction model was developed without any external validation</u> , and it was rated as <u>low risk of bias for all domains</u> , consider downgrading to high risk of bias . Such a model can only be considered as low risk of bias, if the development was based on a very large data set and included some form of internal validation.
High risk of bias	If at least one domain is judged to be at high risk of bias .
Unclear risk of bias	If an unclear risk of bias was noted in at least one domain and it was low risk for all other domains.

Complete for each evaluation of a distinct model.

Reaching an overall judgment about applicability of the prediction model evaluation	
Low concerns for applicability	If low concerns for applicability for all domains, the prediction model evaluation is judged to have low concerns for applicability .
High concerns for applicability	If high concerns for applicability for at least one domain, the prediction model evaluation is judged to have high concerns for applicability .

Unclear concerns for applicability	If unclear concerns (but no “high concern”) for applicability for at least one domain, the prediction model evaluation is judged to have unclear concerns for applicability overall.	
Overall judgment about risk of bias and applicability of the prediction model evaluation		
Overall judgment of risk of bias	RISK: (low/ high/ unclear)	Unclear
Summary of sources of potential bias: Unclear		
Overall judgment of applicability	CONCERN: (low/ high/ unclear)	Low
Summary of applicability concerns: Low		

PROBAST = Prediction model Risk Of Bias ASsessment Tool.

16. Jacob J, et al:

Step 2: Classify the type of prediction model evaluation

Classify the evaluation based on its aim			
Type of prediction study	PROBAST boxes to complete	Tick as appropriate	Definitions for type of prediction model study
Development only	Dev		Prediction model development without external validation. These studies may include internal validation methods such as bootstrapping and cross-validation techniques
Development and validation	Dev and Val	X	Prediction model development combined with external validation in other participants in the same article
Validation only	Val		External validation of existing (previously developed) model in other participants
<i>This table should be completed once for each publication being assessed and for each relevant outcome in your review.</i>			
Publication reference	Jacob J, Miró Ò, Herrero P, Martín-Sánchez FJ, Gil V, Tost J, et al. Predicción de la mortalidad a muy corto plazo de los pacientes con insuficiencia cardiaca crónica agudizada: escala EAHFE-3D. Med Intensiva. 2016; 40:348---355.		
Models of interest	Escala EAHFE-3D		
Outcome of interest	The outcome was to predict mortality at 3 days		

Step 3: Assess risk of bias and applicability

DOMAIN 1: Participants			
A. Risk of Bias			
<p><i>Describe the sources of data and criteria for participant selection: The EAHFE-3D study is a secondary analysis of the Epidemiology of Acute Heart Failure in Emergency Departments (EAHFE) registry. It is a multipurpose cohort study of an analytical, non-interventional and multicenter nature with prospective follow-up that consecutively included all patients seen due to ECHF in 34 Spanish ECDs. The patient inclusion criteria were compliance with the diagnostic criteria for ECHF based on the presence of symptoms (dyspnea, orthopnea, paroxysmal nocturnal dyspnea) and signs (third sound, lung crepitations, jugular venous pressure > 4 cm, resting sinus node tachycardia, edemas, hepatomegalia, hepatojugular reflux), and radiological evidence of pulmonary congestion, requiring immediate treatment for stabilization. The EAHFE study has had four patient enrollment phases. In order to generate the prognostic scale, use was made of the data collected in EAHFE-1 (from 15 April to 15 May 2007, with 1107 patients), EAHFE-2 (from 1 to 30 June 2009, with 1483 patients) and EAHFE-3 (from 7 November 2011 to 7 January 2012, with 3255 patients)-representing a total of 5845 patients enrolled in 29 ECDs. For the present post hoc study we included those patients with all the values needed to perform the intended analysis and corresponding to the follow-up period of three days. In order to validate the scale we used the data of the EAHFE-4 (from 1 January to 28 February 2014, with 2957 patients enrolled in 34 ECDs). The data collection methodology was the same in all four periods and in all centers, and has been published elsewhere. The study was approved by the Clinical Research Ethics Committees of the participating hospitals, and informed consent was obtained from the patients included in the study.</i></p>			
		De v	Val
1.1 Were appropriate data sources used, e.g. cohort, RCT or nested case-control study data?		Y	Y
1.2 Were all inclusions and exclusions of participants appropriate?		Y	Y
Risk of bias introduced by selection of participants	RISK: (low/ high/ unclear)	Low	Low
<i>Rationale of bias rating: Applied correct criteria for inclusion of participants</i>			
B. Applicability			
<p><i>Describe included participants, setting and dates: Of the 5845 episodes included in the EAHFE registry, a total of 3640 for which 100% of the information was available for the derivation cohort were analyzed. Of these patients, 101 (2.8%) died in the course of the three days after ECD consultation. In the derivation cohort, of the 102 patients that had died after three days, 16.7% had been discharged home (n = 17). Of these, 70.6% (n = 12) again visited the ECD; as a result, 95.1% finally died in hospital (n = 97). Regarding the destination of the patients that died, 8.9% (n = 9) were considered amenable to intensive care and were admitted to the ICU. Of note in this regard is that up to 66% of the patients (n = 57) were admitted to units dependent upon Internal Medicine.</i></p>			
Concern that the included participants and setting do not match the review question	CONCERN: (low/ high/ unclear)	Low	Low
<i>Rationale of applicability rating: The characteristics of the study population are consistent with the criteria of the research question.</i>			

DOMAIN 2: Predictors	
A. Risk of Bias	
<p><i>List and describe predictors included in the final model, e.g. definition and timing of assessment: Age ≥75 years, Baseline NYHA functional class III-IV, Sat O₂<90% Hyponatremia (<135 mmol/l), Use of inotropic or vasopressor drugs, Use of noninvasive ventilation, Systolic blood pressure < 110 mmHg. The cutoff points chosen for each of the 5 early mortality risk groups were: very low risk group 0---20 points; low risk group 21-40 points; intermediate risk group 41-60 points; high risk group 61-80 points; and very high risk group > 80 points.</i></p>	

	De v	Val
2.1 Were predictors defined and assessed in a similar way for all participants?	Y	Y
2.2 Were predictor assessments made without knowledge of outcome data?	Y	Y
2.3 Are all predictors available at the time the model is intended to be used?	Y	Y
Risk of bias introduced by predictors or their assessment	RISK: (low/ high/ unclear)	Low Low
<i>Rationale of bias rating: It has a complete description of the list of selected predictor variables, as well as the way in which they were studied in patients.</i>		
B. Applicability		
Concern that the definition, assessment or timing of predictors in the model do not match the review question	CONCERN: (low/ high/ unclear)	Low Low
<i>Rationale of applicability rating: The evaluation time of the predictive variables and the applicability of these are consistent with the research question of the recent systematic review.</i>		
DOMAIN 3: Outcome		
A. Risk of Bias		
<i>Describe the outcome, how it was defined and determined, and the time interval between predictor assessment and outcome determination: The outcome is to predict mortality in a 3 day range.</i>		
	Dev	Val
3.1 Was the outcome determined appropriately?	Y	Y
3.2 Was a pre-specified or standard outcome definition used?	Y	Y
3.3 Were predictors excluded from the outcome definition?	Y	Y
3.4 Was the outcome defined and determined in a similar way for all participants?	Y	Y
3.5 Was the outcome determined without knowledge of predictor information?	Y	Y
3.6 Was the time interval between predictor assessment and outcome determination appropriate?	Y	Y
Risk of bias introduced by the outcome or its determination	RISK: (low/ high/ unclear)	Low Low
<i>Rationale of bias rating: It has an adequate outcome definition that matches the research question in our review.</i>		
B. Applicability		
<i>At what time point was the outcome determined: It was determined during the follow-up in the future, during the time of hospitalization.</i>		
<i>If a composite outcome was used, describe the relative frequency/distribution of each contributing outcome:</i>		
Concern that the outcome, its definition, timing or determination do not match the review question	CONCERN: (low/ high/ unclear)	Low Low
<i>Rationale of applicability rating: The time in which the outcome is determined is consistent with the proposed research question.</i>		
DOMAIN 4: Analysis		
Risk of Bias		

<p><i>Describe numbers of participants, number of candidate predictors (for DEV only), outcome events and events per candidate predictor (for DEV only):</i> Of the 5845 episodes included in the EAHFE registry, a total of 3640 for which 100% of the information was available for the derivation cohort were analyzed. Of these patients, 101 (2.8%) died in the course of the three days after ECD consultation. In the derivation cohort, of the 102 patients that had died after three days, 16.7% had been discharged home (n = 17). Of these, 70.6% (n = 12) again visited the ECD; as a result, 95.1% finally died in hospital (n = 97). Regarding the destination of the patients that died, 8.9% (n = 9) were considered amenable to intensive care and were admitted to the ICU. Of note in this regard is that up to 66% of the patients (n = 57) were admitted to units dependent upon Internal Medicine. they do not mention it explicitly, but probably the number of candidate variables was 13, with a number of events of 102. EPV: 7.84.</p>
<p><i>Describe how the model was developed (predictor selection, optimism, risk groups, model performance):</i> Adjustment of the multivariate model showed statistical significance to be retained for age ≥ 75 years (OR 3.1; 95%CI 1.5 - 6.3; $p < 0.001$), baseline NYHA functional class iii-iv (OR 1.6; 95%CI 1.1- 2.5; $p = 0.03$), SBP < 110 mmHg (OR 2.2; 95%CI 1.3---3.8; $p = 0.002$), SatO₂ $< 90\%$ (OR 3.2; 95%CI 2.1 - 5; $p < 0.001$), hyponatremia (OR 1.9; 95%CI 1.2 - 2.9; $p = 0.005$), inotropic or vasopressor treatment (OR 2.9; 95%CI 1.5 - 5.7; $p = 0.002$), and the need for NIV (OR 2; 95%CI 1.2 - 3.5; $p = 0.01$). The cutoff points chosen for each of the 5 early mortality risk groups were: very low risk group 0- 20 points; low risk group 21- 40 points; intermediate risk group 41- 60 points; high risk group 61---80 points; and very high risk group > 80 points. Qualitative variables were reported as absolute and relative frequencies, while quantitative variables were reported as the mean and standard deviation. Comparisons were made using the chi-squared test (or 2×2 tables with the Fisher exact test when the expected values were under 5) in the case of qualitative variables, and the Student t-test for independent measures in the case of quantitative variables. Those variables found to present significant differences in the univariate analysis, expressed as odds ratios (ORs), were entered in a logistic regression model. The variables that maintained statistical significance in the multivariate analysis were used to create a weighted scale according to their impact in the final model. The discrimination capacity of the model was assessed based on the area (AUC) under the receiver operating characteristic (ROC) curve. An analysis was subsequently made of the distribution curve of the EAHFE-3D scores according to mortality, with the arbitrary definition of 5 risk groups according to those cutoff points that made clinical sense. These cutoff points defined ascending risk categories (very low, low, intermediate, and very high risk, respectively). Differences were considered statistically significant for $p < 0.05$, or when the 95% confidence interval (CI) of the OR excluded the value 1, or when the AUC excluded the value 0.5. Validation was made with the EAHFE-4 cohort, and its AUC was calculated. Comparison of the AUC was carried out using the methodology of DeLong. The SPSS® version 19.0 statistical package was used, and the program STATA® 12.0 was employed for the comparison of AUC.</p>
<p><i>Describe whether and how the model was validated, either internally (e.g. bootstrapping, cross validation, random split sample) or externally (e.g. temporal validation, geographical validation, different setting, different type of participants):</i> The article mentions the use of derivation and validation cohorts, for the internal and external process.</p>
<p><i>Describe the performance measures of the model, e.g. (re)calibration, discrimination, (re)classification, net benefit:</i> highThe number of patients included in each group was 432 (11.9%), 1329 (36.5%), 1080 (29.7%), 579 (15.9%) and 220 (6.0%), respectively, with a recorded mortality rate of 0, 0.8, 2.9, 5.5 and 12.7%. The discriminatory capacity of the model was good, with an AUC of 0.80 (95%CI 0.76---0.84; $p < 0.001$). The validation cohort in turn included 2957 patients, of which 66 died after three days (2.2%). The AUC of this model was 0.76 (95%CI 0.70 -0.82; $p < 0.001$) (Fig. 4) - the mortality rate by groups being 0.5, 1.0, 2.8, 5.8 and 22.4, respectively. The difference between the AUC was 0.04. The DeLong nonparametric method showed no significant differences in the comparison of AUC ($p = 0.47$).</p>
<p><i>Describe any participants who were excluded from the analysis:</i> This study only excluded patients with STsegment elevation acute coronary syndrome. They don't provide data about it.</p>
<p><i>Describe missing data on predictors and outcomes as well as methods used for missing data:</i> Doesn't describe what was done with the lost data</p>

		Dev	Val
4.1	Were there a reasonable number of participants with the outcome?	N	N
4.2	Were continuous and categorical predictors handled appropriately?	PY	PY
4.3	Were all enrolled participants included in the analysis?	Y	Y
4.4	Were participants with missing data handled appropriately?	NI	NI
4.5	Was selection of predictors based on univariable analysis avoided?	Y	Y
4.6	Were complexities in the data (e.g. censoring, competing risks, sampling of controls) accounted for appropriately?	Y	Y
4.7	Were relevant model performance measures evaluated appropriately?	PY	PY
4.8	Were model overfitting and optimism in model performance accounted for?	NI	NI
4.9	Do predictors and their assigned weights in the final model correspond to the results from the reported multivariable analysis?	Y	Y
Risk of bias introduced by the analysis		RISK: (low/ high/ unclear)	High High
<i>Rationale of bias rating: There are missing data on the candidate variables and the way in which they were chosen, in addition to not mentioning the handling of the missing data. They incompletely evaluated the performance, since the model is missing calibration.</i>			

Step 4: Overall assessment

Use the following tables to reach overall judgments about risk of bias and concerns for applicability of the prediction model evaluation (development and/or validation) across all assessed domains.

Reaching an overall judgment about risk of bias of the prediction model evaluation	
Low risk of bias	If all domains were rated low risk of bias. If a prediction model was developed without any external validation, and it was rated as <u>low risk of bias for all domains</u> , consider downgrading to high risk of bias . Such a model can only be considered as low risk of bias, if the development was based on a very large data set and included some form of internal validation.
High risk of bias	If at least one domain is judged to be at high risk of bias .
Unclear risk of bias	If an unclear risk of bias was noted in at least one domain and it was low risk for all other domains.

Complete for each evaluation of a distinct model.

Reaching an overall judgment about applicability of the prediction model evaluation	
Low concerns for applicability	If low concerns for applicability for all domains, the prediction model evaluation is judged to have low concerns for applicability .
High concerns for applicability	If high concerns for applicability for at least one domain, the prediction model evaluation is judged to have high concerns for applicability .
Unclear concerns for applicability	If unclear concerns (but no "high concern") for applicability for at least one domain, the prediction model evaluation is judged to have unclear concerns for applicability overall.

Overall judgment about risk of bias and applicability of the prediction model evaluation		
Overall judgment of risk of bias	RISK: (low/ high/ unclear)	
Summary of sources of potential bias: High		
Overall judgment of applicability	CONCERN: (low/ high/ unclear)	
Summary of applicability concerns: Low		

PROBAST = Prediction model Risk Of Bias ASsessment Tool.

17. Jia Q, et al.

Step 2: Classify the type of prediction model evaluation

Classify the evaluation based on its aim			
Type of prediction study	PROBAST boxes to complete	Tick as appropriate	Definitions for type of prediction model study
Development only	Dev		Prediction model development without external validation. These studies may include internal validation methods such as bootstrapping and cross-validation techniques
Development and validation	Dev and Val	X	Prediction model development combined with external validation in other participants in the same article
Validation only	Val		External validation of existing (previously developed) model in other participants
<i>This table should be completed once for each publication being assessed and for each relevant outcome in your review.</i>			
Publication reference	Qian JI, Yu-Rong W, Ping HE, Xue-Liang H, Wei Y, Yang MU, et al. Prediction model of in-hospital mortality in elderly patients with acute heart failure based on retrospective study. J Geriatr Cardiol 2017; 14: 669-678. doi:10.11909/j.issn.1671-5411.2017.11.002		
Models of interest	Prediction model of in-hospital mortality in elderly patients with acute heart failure based on retrospective study.		
Outcome of interest	Predictive of in-hospital mortality in elderly patients hospitalized with acute heart failure (AHF).		

Step 3: Assess risk of bias and applicability

DOMAIN 1: Participants			
A. Risk of Bias			
Describe the sources of data and criteria for participant selection: This retrospective study was performed in intensive care units of the Cardiology Department in the Chinese PLA General Hospital. For the purpose of our study, diagnosis of AHF was made according to the inclusion criteria from the European Society of Cardiology according to symptoms or signs, electrocardiogram, chest radiograph, and echocardiography. Patients were identified for inclusion in the study from admission when given a diagnosis of AHF based on International Classification of Diseases, Ninth/Tenth Revision (ICD-9/10) coding from 2003 January to 2016 October (n = 3026).			
		De v	Val
1.1 Were appropriate data sources used, e.g. cohort, RCT or nested case-control study data?		Y	Y
1.2 Were all inclusions and exclusions of participants appropriate?		Y	Y
Risk of bias introduced by selection of participants	RISK: (low/ high/ unclear)	Low	Low
Rationale of bias rating: Applied correct criteria for inclusion of participants			
B. Applicability			
Describe included participants, setting and dates: In total, 2431 patients with AHF were included in this retrospective study. Based on the grouping method of a previous study, 2431 patients were randomly divided into derivation (70%, n = 1702) and validation (30%, n = 729) groups.[9] The baseline characteristics, physiological parameters and management of patients in the derivation group are shown in Table 1. A total of 1702 patients were included. They were divided into two groups according to their outcomes during hospitalization. In total, 90 patients who died and 1612 who survived during hospitalization were included. The two groups were similar with respect to age, sex distribution, systolic blood pressure, diastolic blood pressure and body mass index. In the past medical history, patients dying during hospital stay had a great number of pulmonary infection (57.77% vs. 37.47%, P < 0.001) compared with patients surviving during hospitalization.			
Concern that the included participants and setting do not match the review question	CONCERN: (low/ high/ unclear)	Low	Low
Rationale of applicability rating: The characteristics of the study population are consistent with the criteria of the research question.			

DOMAIN 2: Predictors			
A. Risk of Bias			
List and describe predictors included in the final model, e.g. definition and timing of assessment: estimated glomerular filtration rate, heart rate, left ventricular ejection fraction, N-terminal pro-b-type natriuretic peptide, pH value			
		De v	Val
2.1 Were predictors defined and assessed in a similar way for all participants?		Y	Y
2.2 Were predictor assessments made without knowledge of outcome data?		Y	Y
2.3 Are all predictors available at the time the model is intended to be used?		Y	Y
Risk of bias introduced by predictors or their assessment	RISK: (low/ high/ unclear)	Low	Low
Rationale of bias rating: It has a complete description of the list of selected predictor variables, as well as the way in which they were studied in patients.			
B. Applicability			

Concern that the definition, assessment or timing of predictors in the model do not match the review question	CONCERN: (low/ high/ unclear)	Low	Low
<i>Rationale of applicability rating:</i> <i>The evaluation time of the predictive variables and the applicability of these are consistent with the research question of the recent systematic review.</i>			
DOMAIN 3: Outcome			
A. Risk of Bias			
<i>Describe the outcome, how it was defined and determined, and the time interval between predictor assessment and outcome determination:</i> Predictive of in-hospital mortality in elderly patients hospitalized with acute heart failure (AHF).			
		Dev	Val
3.1 Was the outcome determined appropriately?		PY	PY
3.2 Was a pre-specified or standard outcome definition used?		Y	Y
3.3 Were predictors excluded from the outcome definition?		Y	Y
3.4 Was the outcome defined and determined in a similar way for all participants?		Y	Y
3.5 Was the outcome determined without knowledge of predictor information?		Y	Y
3.6 Was the time interval between predictor assessment and outcome determination appropriate?		PY	PY
Risk of bias introduced by the outcome or its determination	RISK: (low/ high/ unclear)	Unclear	Unclear
<i>Rationale of bias rating: They don't describe data on the follow-up time.</i>			
B. Applicability			
<i>At what time point was the outcome determined:</i> <i>It was determined during the follow-up in the future, during the time of hospitalization.</i> <i>If a composite outcome was used, describe the relative frequency/distribution of each contributing outcome:</i>			
Concern that the outcome, its definition, timing or determination do not match the review question	CONCERN: (low/ high/ unclear)	Low	Low
<i>Rationale of applicability rating: The time in which the outcome is determined is consistent with the proposed research question.</i>			
DOMAIN 4: Analysis			
Risk of Bias			
<i>Describe numbers of participants, number of candidate predictors (for DEV only), outcome events and events per candidate predictor (for DEV only):</i> In total, 2431 patients with AHF were included in this retrospective study. Based on the grouping method of a previous study, 2431 patients were randomly divided into derivation (70%, n = 1702) and validation (30%, n = 729) groups. They don't mention the candidate variables			
<i>Describe how the model was developed (predictor selection, optimism, risk groups, model performance):</i> All statistical operations in this study were completed by the Department of Statistics of Peking University Health Science Center. The normal distribution test for continuous variables was conducted using the Kolmogorov-Smirnov test. Continuous variables were presented as the mean ± SD or median (25th–75th percentiles) according to normality. Statistical analysis of normal distribution data was performed using an unpaired Student's t-test between two groups. Non-normal distribution data were analyzed using the Mann-Whitney U			

<p>test for continuous variables and the x2 test for discrete variables. Multivariate analysis was conducted using binary logistic regression. Statistical results are expressed as odds ratios (OR) and 95% confidence intervals (CI). We used the forward method to exclude insignificant variables. The receiver operator characteristic curve (ROC) of significant variables and C-statistic testing were utilized to assess the performance of the constructed model in comparison with a previously published model. All tests were carried out using SPSS version 17.0 (SPSS Inc., Chicago, IL, USA) and MedCalc version 15.6.1 (MedCalc Software, Mariakerke, Belgium) statistical software. Differences were considered statistically significant at a twotailed P value of less than 0.05.</p>		
<p><i>Describe whether and how the model was validated, either internally (e.g. bootstrapping, cross validation, random split sample) or externally (e.g. temporal validation, geographical validation, different setting, different type of participants):</i> They do not mention internal or external validation methods.</p>		
<p><i>Describe the performance measures of the model, e.g. (re)calibration, discrimination, (re)classification, net benefit:</i> According to the results presented in Table 1, heart rate (HR), respiratory rate (RR), aspartate aminotransferase (AST), eGFR (calculated by MDRD method), NT-pro BNP, pH value, HCO₃⁻, left ventricular diastolic diameter (LVDD), left ventricular systolic diameter (LVSD), LVEF, FS and past history of cardiomyopathy, pulmonary infection, and chronic kidney disease were entered into a multivariable logistic regression model. The significant variables included HR (OR: 1.043, 95% CI: 1.030–1.057, P < 0.001), LVEF (OR: 0.918, 95% CI: 0.873–0.966, P = 0.001), pH value (OR: 0.001, 95% CI: 0.000–0.002, P < 0.001), eGFR (OR: 0.120, 95% CI: 0.066–0.220, P < 0.001), and NT-pro BNP (OR: 3.463, 95% CI: 1.870–6.413, P < 0.001). All patients were stratified according to quartiles of HR (Q1, 43.00 to 75.00; Q2, 76.00 to 83.00; Q3, 84.00 to 103.00; and Q4, 104.00 to 166.00), pH value (Q1, 6.785 to 7.366; Q2, 7.367 to 7.400; Q3, 7.401 to 7.435; and Q4, 7.436 to 7.524), eGFR (Q1, 24.40 to 33.47; Q2, 33.48 to 63.07; Q3, 63.08 to 86.65; and Q4, 86.66 to 120.24) and NT-pro BNP (Q1, 950.5 to 2257.5; Q2, 2257.6 to 5737.5; Q3, 5737.6 to 9369.5; and Q4, 9369.6 to 3,5000). (A–E). The incidence of patients in the deceased group for Q1 was higher than the other three groups according to the pH value and eGFR, while the incidence of the deceased group in Q1 was the lowest according to HR and NT-pro BNP. Furthermore, the incidence of patients dying during hospitalization with reduced LVEF was higher than that of patients with preserved LVEF. Acid-base imbalance often occurs in HF patients, and the pH value is an excellent warning indicator of adverse cardiac events. (A–C) shows the distribution of the pH values for all patients. Using the various indicators of the ROC curve, we calculated Youden's index to determine the cutoff value of each index. We established the following in-hospital mortality prediction algorithm (PA): $\text{logit}(P) = 78.023 + 0.042 \times \text{HR} - 0.085 \times \text{LVEF} - 6.028 \times \text{pH} - 2.118 \times \text{eGFR} + 1.242 \times \text{NT pro-BNP}$. The closer the P value is to 1, the greater the likelihood of adverse cardiac events. To facilitate clinical practice, we used the five independent risk factors selected by binary logistic regression to establish a Risk Prediction Score System (RPSS). According to the OR value of the independent risk factors screened by logistic regression for in-hospital mortality, the risk factors were assigned, the non-integral OR values were taken from four to five, and the corresponding risk integral values of risk factors were obtained. The area under the ROC curve values for RPSS, PA and GWTG-HF were 0.873, 0.829, and 0.818, respectively (Figure 5). The cut-off of the ROC curve for RPSS is four points. Therefore, clinical doctors should focus special attention on patients with more than four points. Furthermore, we compared the characteristics of the derivation group and validation group. There is no significant difference between the two groups. The area under the ROC curve values for the validation RPSS is 0.863. The cut-off of the ROC curve for the validation group is also four points.</p>		
<p><i>Describe any participants who were excluded from the analysis:</i> The exclusion criteria included the following: younger than 60 years old (n = 295), malignant tumors (n = 23), acute coronary syndrome (n = 174), acute cerebrovascular diseases within the past three months (n = 37), immunological diseases (n = 11) and other patients in any case that were not suitable for this study. According to the criteria mentioned above, 2486 patients who were 60 years and older with AHF in intensive care units of the Chinese PLA General Hospital were screened, and 55 patients were excluded for missing data.</p>		
<p><i>Describe missing data on predictors and outcomes as well as methods used for missing data:</i> They don't mention the handling of lost data</p>		
	Dev	Val
4.1 Were there a reasonable number of participants with the outcome?	PN	PN
4.2 Were continuous and categorical predictors handled appropriately?	Y	Y
4.3 Were all enrolled participants included in the analysis?	Y	Y

4.4 Were participants with missing data handled appropriately?		NI	NI
4.5 Was selection of predictors based on univariable analysis avoided?		Y	Y
4.6 Were complexities in the data (e.g. censoring, competing risks, sampling of controls) accounted for appropriately?		Y	Y
4.7 Were relevant model performance measures evaluated appropriately?		PY	PY
4.8 Were model overfitting and optimism in model performance accounted for?		NI	NI
4.9 Do predictors and their assigned weights in the final model correspond to the results from the reported multivariable analysis?		Y	Y
Risk of bias introduced by the analysis	RISK: (low/ high/ unclear)	High	High
<i>Rationale of bias rating: They incompletely evaluated the performance, since the model is missing calibration.</i>			

Step 4: Overall assessment

Use the following tables to reach overall judgments about risk of bias and concerns for applicability of the prediction model evaluation (development and/or validation) across all assessed domains.

Reaching an overall judgment about risk of bias of the prediction model evaluation	
Low risk of bias	If all domains were rated low risk of bias. If a <u>prediction model was developed without any external validation</u> , and it was rated as <u>low risk of bias for all domains</u> , consider downgrading to high risk of bias . Such a model can only be considered as low risk of bias, if the development was based on a very large data set <u>and</u> included some form of internal validation.
High risk of bias	If at least one domain is judged to be at high risk of bias .
Unclear risk of bias	If an unclear risk of bias was noted in at least one domain and it was low risk for all other domains.

Complete for each evaluation of a distinct model.

Reaching an overall judgment about applicability of the prediction model evaluation	
Low concerns for applicability	If low concerns for applicability for all domains, the prediction model evaluation is judged to have low concerns for applicability .
High concerns for applicability	If high concerns for applicability for at least one domain, the prediction model evaluation is judged to have high concerns for applicability .
Unclear concerns for applicability	If unclear concerns (but no “high concern”) for applicability for at least one domain, the prediction model evaluation is judged to have unclear concerns for applicability overall.

Overall judgment about risk of bias and applicability of the prediction model evaluation		
Overall judgment of risk of bias	RISK: (low/ high/ unclear)	

Summary of sources of potential bias: High		
Overall judgment of applicability	CONCERN: (low/ high/ unclear)	
Summary of applicability concerns: Low		

PROBAST = Prediction model Risk Of Bias ASsessment Tool.

18. Khanam SS, et al:

Step 2: Classify the type of prediction model evaluation

Classify the evaluation based on its aim			
Type of prediction study	PROBAST boxes to complete	Tick as appropriate	Definitions for type of prediction model study
Development only	Dev		Prediction model development without external validation. These studies may include internal validation methods such as bootstrapping and cross-validation techniques
Development and validation	Dev and Val	X	Prediction model development combined with external validation in other participants in the same article
Validation only	Val		External validation of existing (previously developed) model in other participants
<i>This table should be completed once for each publication being assessed and for each relevant outcome in your review.</i>			
Publication reference	Khanam SS, Choi E, Son J-W, Lee J-W, Youn YJ, Yoon J, et al. (2018) Validation of the MAGGIC (Meta-Analysis Global Group in Chronic Heart Failure) heart failure risk score and the effect of adding natriuretic peptide for predicting mortality after discharge in hospitalized patients with heart failure. PLoS ONE 13(11): e0206380. https://doi.org/10.1371/journal.pone.0206380		
Models of interest	MAGGIC risk score with BNP or NT-proBNP.		
Outcome of interest	Derive and externally validate a risk score to predict 1-year mortality after discharge in hospitalized patients with HF using the MAGGIC program data. We also assessed the effect of adding BNP or NT-proBNP to this risk score model in a Korean heart failure registry population.		

Step 3: Assess risk of bias and applicability

DOMAIN 1: Participants
A. Risk of Bias

Describe the sources of data and criteria for participant selection: The Korean Acute Heart Failure (KorAHF) registry is a prospective multicenter cohort study designed to measure the outcomes of Korean patients admitted for HF. The registry data consist of patient demographics, clinical characteristics, and evidence-based treatments. Patients who had signs or symptoms of HF and met one of the following criteria were eligible for this study: (1) lung congestion or (2) objective findings of left ventricular systolic dysfunction or structural heart disease. Patients hospitalized for AHF from 10 tertiary university hospitals throughout the country were consecutively enrolled from March 2011 to February 2014. The patients were planned to be followed up until 2018. Data were collected by each site and entered into a web based case-report form in the Clinical Data Management System from the Korea National Institute of Health.

	De v	Val
1.1 Were appropriate data sources used, e.g. cohort, RCT or nested case-control study data?		Y
1.2 Were all inclusions and exclusions of participants appropriate?		Y
Risk of bias introduced by selection of participants	RISK: (low/ high/ unclear)	Low

Rationale of bias rating:

The selection criteria are mentioned in the study

B. Applicability

Describe included participants, setting and dates: We enrolled 5,625 AHF subjects from 10 tertiary university hospitals in Korea. We excluded 269 patients who died in hospital, 784 patients with missing patient characteristics, and patients without death information. Finally, we analyzed 4572 patients including 1623 (35.5%) dead and 2949 (64.5%) alive. The mean age was 68.42 ± 14.52 years, and 2392 (52.32%) were men. The overall 1-year mortality in the study population was 17.58%. Among the 13 variables, age, creatinine (Cr), smoker and diabetes (41.53%), COPD (14.17%), HF duration > 18 month (52.37%) and NYHA 4 (53.73%) as well as not taking medications (ACEI or ARB and beta blocker) were significantly higher in patients who died during follow-up compared with those who are alive. Additionally, body mass index (BMI) was almost similar between the 2 groups.

Concern that the included participants and setting do not match the review question	CONCERN: (low/ high/ unclear)	Low
--	---	-----

Rationale of applicability rating: Information from patients were enrolled in at admission from 2011 to 2014.

DOMAIN 2: Predictors

A. Risk of Bias

List and describe predictors included in the final model, e.g. definition and timing of assessment: 11 variables included in the MAGGIC score were associated with mortality. However, when BNP or NT-pro-BNP were associated with MAGGIC. ejection fraction could not predict mortality (age, male sex, BMI, SBP, diabetes, HF duration > 18 months, NYHA (2–4), Cr, ACEI or ARB, and beta blocker, along with high BNP or NT-proBNP)

	De v	Val
2.1 Were predictors defined and assessed in a similar way for all participants?		Y
2.2 Were predictor assessments made without knowledge of outcome data?		Y
2.3 Are all predictors available at the time the model is intended to be used?		Y
Risk of bias introduced by predictors or their assessment	RISK: (low/ high/ unclear)	Low

Rationale of bias rating: The article added variable to a list of previous demonstrated model.

B. Applicability

Concern that the definition, assessment or timing of predictors in the model do not match the review question	CONCERN: (low/ high/ unclear)		Low
<i>Rationale of applicability rating: They use a simple group of variables, easy to get for develop one usefull and very applicable model.</i>			
DOMAIN 3: Outcome			
A. Risk of Bias			
<i>Describe the outcome, how it was defined and determined, and the time interval between predictor assessment and outcome determination: we aimed to derive and externally validate a risk score to predict 1-year mortality after discharge in hospitalized patients with HF using the MAGGIC program data. We also assessed the effect of adding BNP or NT-proBNP to this risk score model in a Korean heart failure registry population.</i>			
		Dev	Val
3.1 Was the outcome determined appropriately?			Y
3.2 Was a pre-specified or standard outcome definition used?			Y
3.3 Were predictors excluded from the outcome definition?			Y
3.4 Was the outcome defined and determined in a similar way for all participants?			Y
3.5 Was the outcome determined without knowledge of predictor information?			Y
3.6 Was the time interval between predictor assessment and outcome determination appropriate?			Y
Risk of bias introduced by the outcome or its determination	RISK: (low/ high/ unclear)		Low
<i>Rationale of bias rating: It was very well defined outcome, with good risk model.</i>			
B. Applicability			
<i>At what time point was the outcome determined: at one time in the future</i>			
<i>If a composite outcome was used, describe the relative frequency/distribution of each contributing outcome:</i>			
Concern that the outcome, its definition, timing or determination do not match the review question	CONCERN: (low/ high/ unclear)		Low
<i>Rationale of applicability rating: Its definition of outcome fit with the study propoosite.</i>			
DOMAIN 4: Analysis			
Risk of Bias			
<i>Describe numbers of participants, number of candidate predictors (for DEV only), outcome events and events per candidate predictor (for DEV only): 4572 patients included. There are no information of candidate predictors because they just mention the 11 variables from the MAGGIC score.</i>			
<i>Describe how the model was developed (predictor selection, optimism, risk groups, model performance): In multivariable analysis, 11 variable included in the MAGGIC score were associated with mortality.</i>			
<i>Describe whether and how the model was validated, either internally (e.g. bootstrapping, cross validation, random split sample) or externally (e.g. temporal validation, geographical validation, different setting, different type of participants): They validated the score by external validation, in another population previously stablished.</i>			

Describe the performance measures of the model, e.g. (re)calibration, discrimination, (re)classification, net benefit: Good discrimination abilities were found, with a C-index of 0.734 from the risk score model to predict 1-year mortality. Hosmer-Lemeshow goodness of fit test for risk score mortality indicated a good calibration ($p = 0.5559$) by plotting predicted versus observed mortality using 10 groups		
Describe any participants who were excluded from the analysis: We excluded 269 patients who died in hospital, 784 patients with missing patient characteristics, and patients without death information.		
Describe missing data on predictors and outcomes as well as methods used for missing data: Do not present any missing data		
	Dev	Val
4.1 Were there a reasonable number of participants with the outcome?		Y
4.2 Were continuous and categorical predictors handled appropriately?		Y
4.3 Were all enrolled participants included in the analysis?		Y
4.4 Were participants with missing data handled appropriately?		Y
4.5 Was selection of predictors based on univariable analysis avoided?		Y
4.6 Were complexities in the data (e.g. censoring, competing risks, sampling of controls) accounted for appropriately?		Y
4.7 Were relevant model performance measures evaluated appropriately?		Y
4.8 Were model overfitting and optimism in model performance accounted for?		Y
4.9 Do predictors and their assigned weights in the final model correspond to the results from the reported multivariable analysis?		Y
Risk of bias introduced by the analysis	RISK: (low/ high/ unclear)	Low
Rationale of bias rating: Describe how they get the score		

Step 4: Overall assessment

Use the following tables to reach overall judgments about risk of bias and concerns for applicability of the prediction model evaluation (development and/or validation) across all assessed domains.	
Reaching an overall judgment about risk of bias of the prediction model evaluation	
Low risk of bias	If all domains were rated low risk of bias. If a <u>prediction model was developed without any external validation</u> , and it was rated as <u>low risk of bias for all domains</u> , consider downgrading to high risk of bias . Such a model can only be considered as low risk of bias, if the development was based on a very large data set and included some form of internal validation.
High risk of bias	If at least one domain is judged to be at high risk of bias .
Unclear risk of bias	If an unclear risk of bias was noted in at least one domain and it was low risk for all other domains.
Complete for each evaluation of a distinct model.	
Reaching an overall judgment about applicability of the prediction model evaluation	
Low concerns for	If low concerns for applicability for all domains, the prediction model evaluation is judged to have low concerns for applicability .

applicability			
High concerns for applicability	If high concerns for applicability for at least one domain, the prediction model evaluation is judged to have high concerns for applicability .		
Unclear concerns for applicability	If unclear concerns (but no “high concern”) for applicability for at least one domain, the prediction model evaluation is judged to have unclear concerns for applicability overall.		
Overall judgment about risk of bias and applicability of the prediction model evaluation			
Overall judgment of risk of bias	<table> <tr> <td>RISK: (low/ high/ unclear)</td><td>Low</td></tr> </table>	RISK: (low/ high/ unclear)	Low
RISK: (low/ high/ unclear)	Low		
<i>Summary of sources of potential bias: Low</i>			
Overall judgment of applicability	<table> <tr> <td>CONCERN: (low/ high/ unclear)</td><td>Low</td></tr> </table>	CONCERN: (low/ high/ unclear)	Low
CONCERN: (low/ high/ unclear)	Low		
<i>Summary of applicability concerns: Low</i>			

PROBAST = Prediction model Risk Of Bias ASsessment Tool.

19. Kinugasa Y, et al:

Step 2: Classify the type of prediction model evaluation

Classify the evaluation based on its aim			
Type of prediction study	PROBAST boxes to complete	Tick as appropriate	Definitions for type of prediction model study
Development only	Dev	X	Prediction model development without external validation. These studies may include internal validation methods such as bootstrapping and cross-validation techniques
Development and validation	Dev and Val		Prediction model development combined with external validation in other participants in the same article
Validation only	Val		External validation of existing (previously developed) model in other participants
<i>This table should be completed once for each publication being assessed and for each relevant outcome in your review.</i>			
Publication reference	Kinugasa Y, Kato M, Sugihara S, Hirai M, Kotani K, Ishida K, et al. A Simple Risk Score to Predict In-Hospital Death of Elderly Patients With Acute Decompensated Heart Failure Hypoalbuminemia as an Additional Prognostic Factor. Circ J 2009; 73: 2276 – 2281		

Models of interest	A Simple Risk Score to Predict In-Hospital Death of Elderly Patients With Acute Decompensated Heart Failure Hypoalbuminemia as an Additional Prognostic Factor
Outcome of interest	The primary outcome was defined as all-cause in-hospital mortality.

Step 3: Assess risk of bias and applicability

DOMAIN 1: Participants			
A. Risk of Bias			
Describe the sources of data and criteria for participant selection: This retrospective, observational, and single-center study was conducted between January 2009 and December 2013, enrolling a total of 1,062 consecutive AHF patients admitted to our hospital. A clinical diagnosis of AHF was made based on clinical presentation, echocardiography assessment, and laboratory testing of natriuretic peptides, and was additionally confirmed by clinical records based on Framingham criteria for the diagnosis of heart failure.			
		De v	Val
1.1 Were appropriate data sources used, e.g. cohort, RCT or nested case-control study data?		Y	
1.2 Were all inclusions and exclusions of participants appropriate?		Y	
Risk of bias introduced by selection of participants	RISK: (low/ high/ unclear)	Low	
Rationale of bias rating: Applied correct criteria for inclusion of participants			
B. Applicability			
Describe included participants, setting and dates: A detailed retrospective chart review was conducted for 349 patients aged 65 years and older who were hospitalized in Tottori University hospital with a primary diagnosis of ADHF from January 2004 to October 2008. The mean age of the overall cohort was 79±6 years, and 56% were male. The prevalence of ischemic heart disease was 28.4%. Systolic dysfunction was relatively less prevalent (34.4%), which suggested that HF with preserved systolic function was more common in this cohort. Hypertension (46.9%) was the most common comorbidity, followed by atrial fibrillation (43.0%). Angiotensin-receptor blockers (ARB) (37.5%) were more likely to be administered to this patient population than angiotensin-converting enzyme inhibitors (ACEI: 26.9%). The β -blockers and spironolactone were administered to 31.8% and 26.9% of patients, respectively. The in-hospital mortality rate was 9.7% (n=34); cardiac: 4.6% (n=16), stroke/vascular: 0.9% (n=3), and noncardiovascular: 4.3% (n=15). The mean and median length of hospital stay was 34±32 days and 25 (interquartile range: 16–41) days, respectively			
Concern that the included participants and setting do not match the review question	CONCERN: (low/ high/ unclear)	Low	
Rationale of applicability rating: The characteristics of the study population are consistent with the criteria of the research question.			

DOMAIN 2: Predictors
A. Risk of Bias

<p>List and describe predictors included in the final model, e.g. definition and timing of assessment: Previous hospitalization for HF, Sodium ≤ 138 mmol/L, Albumin ≤ 3.2 g/dl, BUN ≥ 35 mg/dl, BNP ≥ 980 pg/ml. Medical records were retrospectively reviewed with regard to demography, medical history, etiology of HF, comorbidities, laboratory data at the time of hospital admission, pre-hospital medication, and in-hospital outcome</p>			
		Dev	Val
2.1 Were predictors defined and assessed in a similar way for all participants?		Y	
2.2 Were predictor assessments made without knowledge of outcome data?		Y	
2.3 Are all predictors available at the time the model is intended to be used?		Y	
Risk of bias introduced by predictors or their assessment		RISK: (low/ high/ unclear)	Low
<p>Rationale of bias rating: It has a complete description of the list of selected predictor variables, as well as the way in which they were studied in patients.</p>			
B. Applicability			
Concern that the definition, assessment or timing of predictors in the model do not match the review question		CONCERN: (low/ high/ unclear)	Low
<p>Rationale of applicability rating: The evaluation time of the predictive variables and the applicability of these are consistent with the research question of the recent systematic review.</p>			
DOMAIN 3: Outcome			
A. Risk of Bias			
<p>Describe the outcome, how it was defined and determined, and the time interval between predictor assessment and outcome determination: Risk stratification for elderly patients with acute decompensated heart failure (ADHF)</p>			
		Dev	Val
3.1 Was the outcome determined appropriately?		PY	
3.2 Was a pre-specified or standard outcome definition used?		Y	
3.3 Were predictors excluded from the outcome definition?		Y	
3.4 Was the outcome defined and determined in a similar way for all participants?		Y	
3.5 Was the outcome determined without knowledge of predictor information?		Y	
3.6 Was the time interval between predictor assessment and outcome determination appropriate?		PY	
Risk of bias introduced by the outcome or its determination		RISK: (low/ high/ unclear)	Unclear
<p>Rationale of bias rating: They don't describe data on the follow-up time.</p>			
B. Applicability			
<p>At what time point was the outcome determined: It was determined during the follow-up in the future, during the time of hospitalization.</p>			
<p>If a composite outcome was used, describe the relative frequency/distribution of each contributing outcome:</p>			

Concern that the outcome, its definition, timing or determination do not match the review question	CONCERN: (low/ high/ unclear)	Low	
<i>Rationale of applicability rating: The time in which the outcome is determined is consistent with the proposed research question.</i>			
DOMAIN 4: Analysis			
Risk of Bias			
<i>Describe numbers of participants, number of candidate predictors (for DEV only), outcome events and events per candidate predictor (for DEV only): A detailed retrospective chart review was conducted for 349 patients aged 65 years and older who were hospitalized in Tottori University hospital with a primary diagnosis of ADHF from January 2004 to October 2008. The study presented 34 events, and 11 candidate variables. EPV: 3.09.</i>			
<i>Describe how the model was developed (predictor selection, optimism, risk groups, model performance): The mean length of hospital stay in the non-survival group was significantly longer compared with the survival group. Continuous variables are expressed as mean $\pm$ SD, whereas categorical variables are expressed as percentages. Comparisons of 2 groups were made using Student's t-test for continuous variables and the χ^2 test for categorical variables. We performed forward, stepwise multivariate logistic regression analyses to determine the independent predictors of in-hospital mortality. We entered age, sex, and all clinical variables associated with in-hospital mortality in the univariate analysis ($P < 0.05$: a strict level was applied because of the relatively small samples of in-hospital mortality) into the multivariate model. Finally, a clinical risk score was created using identified independent predictors of in-hospital mortality that were dichotomized at cutoff points. The cutoff point for each variable was determined by receiver-operator characteristic (ROC) curves. Each predictor was assigned a number of points proportional to its regression coefficient, as previously described. A risk score was assigned to each subject by adding up the points for each risk factor present. Subjects were then divided into 3 groups according to their risk scores. The area under the ROC curves (AUC) was used to assess the performance of the model in predicting in-hospital mortality. A P value < 0.05 was considered statistically significant. All analyses were performed using SPSS version 11.5 J (SPSS, Inc, Chicago, IL, USA).</i>			
<i>Describe whether and how the model was validated, either internally (e.g. bootstrapping, cross validation, random split sample) or externally (e.g. temporal validation, geographical validation, different setting, different type of participants): They don't mention internal validation methods.</i>			
<i>Describe the performance measures of the model, e.g. (re)calibration, discrimination, (re)classification, net benefit: Multivariate logistic regression analysis identified previous hospitalization for HF, lower serum sodium and albumin concentrations, and elevated BUN and BNP concentrations as independent predictors of in-hospital mortality. Lower systolic blood pressure tended to be associated with in-hospital mortality, but did not reach statistical significance ($P = 0.06$). A risk score was developed by using these variables, which were dichotomized at cutoff points derived from the ROC curves. Each predictor was then assigned a number of points proportional to its regression coefficient as follows: previous hospitalization for HF (2 points), sodium ≤ 138 mmol/L (2 points), albumin ≤ 3.2 g/dl (3 points), BUN ≥ 35 mg/dl (2 points), and BNP ≥ 980 pg/ml (2 points). The regression coefficient of albumin was approximately 1.5, whereas for the other variables it was approximately 1. Thus we assigned 3 points to albumin, and 2 points to the other variable. The AUC for the risk score model was 0.86. We calculated a risk score for each patient by adding the points of each risk factor and stratified the patients into 3 groups: low risk (0–4 points; $n = 210$), moderate risk (5–7 points; $n = 108$), and high risk (8–11 points; $n = 31$). The respective in hospital mortality rates were 1.6%, 15.8%, and 42.1% ($P < 0.05$).</i>			
<i>Describe any participants who were excluded from the analysis: They doesn't mention the characteristics of the excluded population.</i>			
<i>Describe missing data on predictors and outcomes as well as methods used for missing data: They don't mention the handling of lost data</i>			
		Dev	Val

4.1 Were there a reasonable number of participants with the outcome?	N	
4.2 Were continuous and categorical predictors handled appropriately?	Y	
4.3 Were all enrolled participants included in the analysis?	N	
4.4 Were participants with missing data handled appropriately?	NI	
4.5 Was selection of predictors based on univariable analysis avoided?	Y	
4.6 Were complexities in the data (e.g. censoring, competing risks, sampling of controls) accounted for appropriately?	Y	
4.7 Were relevant model performance measures evaluated appropriately?	PY	
4.8 Were model overfitting and optimism in model performance accounted for?	NI	
4.9 Do predictors and their assigned weights in the final model correspond to the results from the reported multivariable analysis?	Y	
Risk of bias introduced by the analysis	RISK: (low/ high/ unclear)	High
<i>Rationale of bias rating: They incompletely evaluated the performance, since the model is missing calibration. They don't mention the handling of missing data.</i>		

Step 4: Overall assessment

Use the following tables to reach overall judgments about risk of bias and concerns for applicability of the prediction model evaluation (development and/or validation) across all assessed domains.

Reaching an overall judgment about risk of bias of the prediction model evaluation	
Low risk of bias	If all domains were rated low risk of bias. If a <u>prediction model was developed without any external validation</u> , and it was rated as <u>low risk of bias for all domains</u> , consider downgrading to high risk of bias . Such a model can only be considered as low risk of bias, if the development was based on a very large data set and included some form of internal validation.
High risk of bias	If at least one domain is judged to be at high risk of bias .
Unclear risk of bias	If an unclear risk of bias was noted in at least one domain and it was low risk for all other domains.

Complete for each evaluation of a distinct model.

Reaching an overall judgment about applicability of the prediction model evaluation	
Low concerns for applicability	If low concerns for applicability for all domains, the prediction model evaluation is judged to have low concerns for applicability .
High concerns for applicability	If high concerns for applicability for at least one domain, the prediction model evaluation is judged to have high concerns for applicability .
Unclear concerns for applicability	If unclear concerns (but no "high concern") for applicability for at least one domain, the prediction model evaluation is judged to have unclear concerns for applicability overall.

Overall judgment about risk of bias and applicability of the prediction model evaluation

Overall judgment of risk of bias	RISK: (low/ high/ unclear)	
<i>Summary of sources of potential bias: Low</i>		
Overall judgment of applicability	CONCERN: (low/ high/ unclear)	
<i>Summary of applicability concerns: Low</i>		

PROBAST = Prediction model Risk Of Bias ASsessment Tool.

20. Lee DS, et al:

Step 2: Classify the type of prediction model evaluation

Classify the evaluation based on its aim			
Type of prediction study	PROBAST boxes to complete	Tick as appropriate	Definitions for type of prediction model study
Development only	Dev		Prediction model development without external validation. These studies may include internal validation methods such as bootstrapping and cross-validation techniques
Development and validation	Dev and Val	X	Prediction model development combined with external validation in other participants in the same article
Validation only	Val		External validation of existing (previously developed) model in other participants
<i>This table should be completed once for each publication being assessed and for each relevant outcome in your review.</i>			
Publication reference	Lee DS, Stitt A, Austin PC, Stukel TA, Schull MJ, Chong A, et al. Prediction of Heart Failure Mortality in Emergent Care: A Cohort Study. Ann Intern Med. 2012;156:767-775.		
Models of interest	The Emergency Heart Failure Mortality Risk Grade (EHMRG).		
Outcome of interest	The primary outcome was 7-day mortality after initial presentation, which is an end point that is important from the perspective of ED decision making and temporally close to the acute care episode.		

Step 3: Assess risk of bias and applicability

DOMAIN 1: Participants
A. Risk of Bias

Describe the sources of data and criteria for participant selection: We studied the data of approximately 12 500 patients who visited an ED for HF and were either discharged or hospitalized from 1 April 2004 to 31 March 2007, in Ontario, Canada, which has a provincial population of more than 13 million. We randomly selected patients, who were either discharged from the ED or hospitalized, for detailed chart abstraction by using stratified cluster sampling. The patients were from hospitals randomly selected from strata defined by hospital type. Approximately 125 hospitalized patients were selected for chart abstraction from each of the 86 hospitals. Patients who were not hospitalized were sampled from a random subsample of 52 hospitals with 50 or more HF visits per year, maintaining a representative proportion of hospital type in the province (that is, teaching, large, or small). Approximately 95 patients were randomly selected from each teaching hospital and large community hospital; similarly, 60 patients were sampled from each small hospital. Heart failure cases were initially identified using the International Classification of Diseases, Tenth Revision (code I50), from the National Ambulatory Care Reporting System for ED visits and the Canadian Institute for Health Information database for patients who were hospitalized. To be included in the study and to emulate ED-based HF diagnosis based on clinical presentation, patients were also required to fulfill the Framingham criteria for HF after examination of their clinical chart records. Patients who were palliative or had do-not- resuscitate orders before ED arrival and transfers from another acute care hospital were excluded. Those who were dialysis-dependent were also excluded because of their differences

in HF pathophysiology and treatment. To assess the performance of the prediction model in a separate cohort, we randomly split the overall sample into derivation and validation data sets. We derived the model on a random sample of approximately 7500 patients with HF, two thirds of whom were hospitalized and one third of whom were discharged from the ED, reflecting the hospitalization-to-discharge ratio from which the study population was sampled. The validation cohort comprised approximately 5000 patients with a similar distribution of hospitalized and discharged patients. Before study initiation, research ethics approval was obtained from Sunnybrook Health Sciences Centre and from hospitals where chart abstraction was done.

	De v	Val
1.1 Were appropriate data sources used, e.g. cohort, RCT or nested case-control study data?	Y	Y
1.2 Were all inclusions and exclusions of participants appropriate?	Y	Y
Risk of bias introduced by selection of participants	RISK: (low/ high/ unclear)	Low Low

Rationale of bias rating: Applied correct criteria for inclusion of participants

B. Applicability

Describe included participants, setting and dates: We examined 15 164 patients with HF who visited the ED and fulfilled the Framingham criteria for HF: 10 781 were hospitalized and 4383 were discharged. Of the 12 591 patients included in the study, 7433 constituted the derivation cohort (5254 hospitalized and 2179 discharged) and 5158 constituted the validation cohort (3560 hospitalized and 1598 discharged). The mean age of the derivation cohort was 75.4 years (SD, 11.4), with 3825 men (51.5%). The mean age of the validation cohort was 75.7 years (SD, 11.4), with 2661 men (51.6%). Overall, 247 deaths occurred, with a 7-day mortality rate of 2.0%. The 7-day mortality rate was 1.8% in the derivation cohort (135 deaths) and 2.2% in the validation cohort (112 deaths). Although baseline characteristics of the 2 cohorts were similar, the validation cohort had marginally higher creatinine concentration, more patients transported by emergency medical services, a greater proportion of patients using diuretics, and more patients with troponin elevation.

Concern that the included participants and setting do not match the review question	CONCERN: (low/ high/ unclear)	Low Low
--	--	---------

Rationale of applicability rating: The characteristics of the study population are consistent with the criteria of the research question.

DOMAIN 2: Predictors

A. Risk of Bias

List and describe predictors included in the final model, e.g. definition and timing of assessment: Age, Transported by EMS, Systolic blood pressure, Heart rate, Oxygen saturation, Creatinine, Potassium, Troponin, Active cancer, Metolazone at home, Adjustment factor. The model can be evaluated during the emergency in a patient with AHF.			
		Dev	Val
2.1 Were predictors defined and assessed in a similar way for all participants?		Y	Y
2.2 Were predictor assessments made without knowledge of outcome data?		Y	Y
2.3 Are all predictors available at the time the model is intended to be used?		Y	Y
Risk of bias introduced by predictors or their assessment	RISK: (low/ high/ unclear)	Low	Low
Rationale of bias rating: They adequately describe the way of choosing final predictors.			
B. Applicability			
Concern that the definition, assessment or timing of predictors in the model do not match the review question	CONCERN: (low/ high/ unclear)	Low	Low
Rationale of applicability rating: The evaluation time of the predictive variables and the applicability of these are consistent with the research question of the recent systematic review.			
DOMAIN 3: Outcome			
A. Risk of Bias			
Describe the outcome, how it was defined and determined, and the time interval between predictor assessment and outcome determination: The primary outcome was 7-day mortality after initial presentation, which is an end point that is important from the perspective of ED decision making and temporally close to the acute care episode.			
		Dev	Val
3.1 Was the outcome determined appropriately?		Y	Y
3.2 Was a pre-specified or standard outcome definition used?		Y	Y
3.3 Were predictors excluded from the outcome definition?		Y	Y
3.4 Was the outcome defined and determined in a similar way for all participants?		Y	Y
3.5 Was the outcome determined without knowledge of predictor information?		Y	Y
3.6 Was the time interval between predictor assessment and outcome determination appropriate?		Y	Y
Risk of bias introduced by the outcome or its determination	RISK: (low/ high/ unclear)	Low	Low
Rationale of bias rating: They show an outcome that matches the question of the systematic review..			
B. Applicability			
At what time point was the outcome determined: It was determined during the follow-up in the future.			
If a composite outcome was used, describe the relative frequency/distribution of each contributing outcome:			
Concern that the outcome, its definition, timing or determination do not match the review question	CONCERN: (low/ high/ unclear)	Low	Low

Rationale of applicability rating: The time in which the outcome is determined is consistent with the proposed research question.

DOMAIN 4: Analysis

Risk of Bias

Describe numbers of participants, number of candidate predictors (for DEV only), outcome events and events per candidate predictor (for DEV only): Of the 12 591 patients included in the study, 7433 constituted the derivation cohort (5254 hospitalized and 2179 discharged) and 5158 constituted the validation cohort (3560 hospitalized and 1598 discharged). The 7-day mortality rate was 1.8% in the derivation cohort (135 deaths) and 2.2% in the validation cohort (112 deaths). With 37 candidate variables, the EPV is: derivation cohort: 3.6 and validation cohort: 3.02

Describe how the model was developed (predictor selection, optimism, risk groups, model performance): Continuous variables were reported as means (SDs) and compared by using a t test or the Kruskal–Wallis test for nonparametric distributions, whereas categorical variables were compared by using the chi-square statistic. In the derivation set, we used univariate logistic regression to examine the association of potential predictors, which are routinely available in the community, with 7-day mortality. Applying methods described by Harrell (18), potential predictors of mortality with P values less than 0.25 were considered for entry into a multiple logistic regression model based on statistical and clinical significance. Covariates were retained in the final multivariate model if the P value was less than 0.05 and if deemed clinically important or previous studies supported their prognostic importance. For continuous variables, the strength and shape of the relationship of potential predictors of death were examined by using cubic spline analyses with 5 knots at percentiles 5, 27.5, 50, 72.5, and 95 of the continuous covariate. On the basis of these analyses, upper and lower bounds were determined for the purposes of identifying truncation values above or below which there was no further contribution to the score calculation. Unmodified β -coefficients were reported from the multiple regression model, and odds ratios (ORs) were reported with estimates greater than unity reflecting increasing mortality risk. Using a method of age-standardized β -coefficient–based weights similar to that used for the Framingham risk score, we developed a scoring system calculated by summing integer scores for categorical variables and weights for the value of continuous variables (where the value of the continuous variable was multiplied by its weight), called the Emergency Heart Failure Mortality Risk Grade (EHMRG). To derive the weights, model β -coefficients were divided by the age coefficient. The resultant weight was then used to determine the contribution of each variable to the overall risk score. Because the weights associated with systolic blood pressure and heart rate were approximately -0.5 and 0.5, respectively, we multiplied all weights by a factor of 2, so that the weight associated with each variable would be an integer. The weights for categorical variables were similarly multiplied by a factor of 2 and rounded to the nearest 5 points.

Describe whether and how the model was validated, either internally (e.g. bootstrapping, cross validation, random split sample) or externally (e.g. temporal validation, geographical validation, different setting, different type of participants):

For the internal validation bootstrapping was used in the derivation cohort, in addition for the external validation a different population was used.

Describe the performance measures of the model, e.g. (re)calibration, discrimination, (re)classification, net benefit: The relationships of age and creatinine concentration with the log odds of 7-day mortality were linear. Systolic blood pressure and oxygen saturation were inversely related to the log odds of death, with an attenuated slope at higher values. Potassium concentration displayed a U-shaped relationship with mortality. In the derivation set, the c-statistic of the multivariate model was 0.805, suggesting high discrimination. The bootstrap-corrected unbiased estimate of the ROC area was 0.811 (95% CI, 0.770 to 0.847). There was no lack of model fit (Hosmer–Lemeshow chi-square statistic, 4.31; $P = 0.828$) and no overfit, as determined by a heuristic linear shrinkage estimator ($\gamma = 0.946$). In the validation data set, the c-statistic of the multivariate model was 0.826, with no lack of model fit (Hosmer–Lemeshow chi-square statistic, 2.99; $P = 0.935$). In the validation data set, the calibration slope was 0.970 (CI, 0.810 to 1.131) and the calibration in the large P value was 0.923, denoting no miscalibration. The Brier score was 0.20. The EHMRG score was normally distributed, with a mean score of 5.9 (SD, 62.0) and a median score of 0.5 (25th, 75th percentiles: -40.2, 44.6). For each 20-point increase in the EHMRG score, the odds of 7-day death increased by 41% (OR, 1.41 [CI, 1.34 to 1.48]) in the derivation cohort and by 39% (OR, 1.39 [CI, 1.32 to 1.47]) in the validation cohort (both $P < 0.001$). For each 1-SD increase in EHMRG, the odds of 7-day mortality were increased 2.9-fold in the derivation (OR, 2.88

[CI, 2.47 to 3.36]) and validation (OR, 2.92 [CI, 2.44 to 3.49]) cohorts (both P < 0.001). The c-statistic of the EHMRG was 0.807 in the derivation cohort and 0.806 (CI, 0.761 to 0.842) after bootstrap correction. The c-statistic was 0.803 in the validation cohort and 0.804 (CI, 0.763 to 0.840) after bootstrap correction.			
<i>Describe any participants who were excluded from the analysis:</i> Patients who were palliative or had do-not-resuscitate orders before ED arrival and transfers from another acute care hospital were excluded. Those who were dialysis-dependent were also excluded because of their differences in HF pathophysiology and treatment. Patients excluded (n = 2573) Palliative: 2246 Dialysis-dependent: 312 Transfers: 15			
<i>Describe missing data on predictors and outcomes as well as methods used for missing data: They don't mention the handling of missing data.</i>			
		Dev	Val
4.1 Were there a reasonable number of participants with the outcome?		Y	Y
4.2 Were continuous and categorical predictors handled appropriately?		Y	Y
4.3 Were all enrolled participants included in the analysis?		N	N
4.4 Were participants with missing data handled appropriately?		NI	NI
4.5 Was selection of predictors based on univariable analysis avoided?		Y	Y
4.6 Were complexities in the data (e.g. censoring, competing risks, sampling of controls) accounted for appropriately?		PY	PY
4.7 Were relevant model performance measures evaluated appropriately?		PY	PY
4.8 Were model overfitting and optimism in model performance accounted for?		NI	NI
4.9 Do predictors and their assigned weights in the final model correspond to the results from the reported multivariable analysis?		Y	Y
Risk of bias introduced by the analysis	RISK: (low/ high/ unclear)	High	High
<i>Rationale of bias rating:</i> The article mentions the measurement of the outcome through calibration and discrimination, however it does not mention the handling of missing data.			

Step 4: Overall assessment

Use the following tables to reach overall judgments about risk of bias and concerns for applicability of the prediction model evaluation (development and/or validation) across all assessed domains.	
Reaching an overall judgment about risk of bias of the prediction model evaluation	
Low risk of bias	If all domains were rated low risk of bias. If a <u>prediction model was developed without any external validation</u> , and it was rated as <u>low risk of bias for all domains</u> , consider downgrading to high risk of bias . Such a model can only be considered as low risk of bias, if the development was based on a very large data set and included some form of internal validation.
High risk of bias	If at least one domain is judged to be at high risk of bias .
Unclear risk of bias	If an unclear risk of bias was noted in at least one domain and it was low risk for all other domains.
<i>Complete for each evaluation of a distinct model.</i>	
Reaching an overall judgment about applicability of the prediction model evaluation	
Low concerns for applicability	If low concerns for applicability for all domains, the prediction model evaluation is judged to have low concerns for applicability .

High concerns for applicability	If high concerns for applicability for at least one domain, the prediction model evaluation is judged to have high concerns for applicability .
Unclear concerns for applicability	If unclear concerns (but no “high concern”) for applicability for at least one domain, the prediction model evaluation is judged to have unclear concerns for applicability overall.
Overall judgment about risk of bias and applicability of the prediction model evaluation	
Overall judgment of risk of bias	RISK: (low/ high/ unclear)
Summary of sources of potential bias: High	
Overall judgment of applicability	CONCERN: (low/ high/ unclear)
Summary of applicability concerns: Low	

PROBAST = Prediction model Risk Of Bias ASsessment Tool.

21. Leong K, et al:

Step 2: Classify the type of prediction model evaluation

Classify the evaluation based on its aim			
Type of prediction study	PROBAST boxes to complete	Tick as appropriate	Definitions for type of prediction model study
Development only	Dev	X	Prediction model development without external validation. These studies may include internal validation methods such as bootstrapping and cross-validation techniques
Development and validation	Dev and Val		Prediction model development combined with external validation in other participants in the same article
Validation only	Val		External validation of existing (previously developed) model in other participants
<i>This table should be completed once for each publication being assessed and for each relevant outcome in your review.</i>			
Publication reference	Leong K. Yin Wong L. Chaw Yu Aung K, Macdonald M, Cao Y. Risk Stratification Model for 30-Day Heart Failure Readmission in a Multiethnic South East Asian Community. <i>The American Journal of Cardiology.</i> (2017). http://dx.doi.org/10.1016/j.amjcard.2017.01.026		
Models of interest	30-day HF readmission risk score		

Outcome of interest	Develop a validated risk score to predict short-term (30-day) HF readmission in patients discharged from hospital with decompensated HF, using readily available and accessible clinical patient information in our hospital.
----------------------------	---

Step 3: Assess risk of bias and applicability

DOMAIN 1: Participants			
A. Risk of Bias			
Describe the sources of data and criteria for participant selection: All patients 21 years or older admitted to the department of cardiology with a principal discharge diagnosis of HF were included in this study. The diagnosis of HF was made if the patient satisfied the clinical Framingham criteria, had an elevated serum N-terminal pro-natriuretic peptide (NT-Pro BNP) level more than the age-specific cut-off level, and evidence of cardiac dysfunction on transthoracic echocardiography, be it systolic or diastolic dysfunction.			
		De v	Val
1.1 Were appropriate data sources used, e.g. cohort, RCT or nested case-control study data?		NI	
1.2 Were all inclusions and exclusions of participants appropriate?		Y	
Risk of bias introduced by selection of participants	RISK: (low/ high/ unclear)	Low	
<i>Rationale of bias rating:</i>			
Inclusion and exclusion criteria were correctly presented.			
B. Applicability			
Describe included participants, setting and dates: The study population consisted of patients with AHF.			
Concern that the included participants and setting do not match the review question	CONCERN: (low/ high/ unclear)	Low	
<i>Rationale of applicability rating: Information from patients were correctly mentioned.</i>			

DOMAIN 2: Predictors			
A. Risk of Bias			
List and describe predictors included in the final model, e.g. definition and timing of assessment: number of previous HF-related admission in the preceding 1 year, index admission length of stay, serum creatinine level, ECG QRS duration, NT-ProBNP level, number of Medical Social Service needs, and b blocker prescription on discharge.			
		De v	Val
2.1 Were predictors defined and assessed in a similar way for all participants?		Y	
2.2 Were predictor assessments made without knowledge of outcome data?		Y	
2.3 Are all predictors available at the time the model is intended to be used?		Y	
Risk of bias introduced by predictors or their assessment	RISK: (low/ high/ unclear)	Low	

<i>Rationale of bias rating: The article has a good selection of variables combining variables known as risk factors in AHF.</i>			
B. Applicability			
Concern that the definition, assessment or timing of predictors in the model do not match the review question	CONCERN: (low/ high/ unclear)	Low	
<i>Rationale of applicability rating: They use a simple group of variables, easy to get for develop one usefull and very applicable model.</i>			
DOMAIN 3: Outcome			
A. Risk of Bias			
Describe the outcome, how it was defined and determined, and the time interval between predictor assessment and outcome determination: Develop a validated risk score to predict short-term (30-day) HF readmission in patients discharged from hospital with decompensated HF, using readily available and accessible clinical patient information in our hospital.			
		Dev	Val
3.1 Was the outcome determined appropriately?		Y	
3.2 Was a pre-specified or standard outcome definition used?		Y	
3.3 Were predictors excluded from the outcome definition?		Y	
3.4 Was the outcome defined and determined in a similar way for all participants?		Y	
3.5 Was the outcome determined without knowledge of predictor information?		Y	
3.6 Was the time interval between predictor assessment and outcome determination appropriate?		Y	
Risk of bias introduced by the outcome or its determination	RISK: (low/ high/ unclear)	Low	
<i>Rationale of bias rating: It was very well defined outcome, with good model</i>			
B. Applicability			
<i>At what time point was the outcome determined: at one time in the future</i>			
<i>If a composite outcome was used, describe the relative frequency/distribution of each contributing outcome:</i>			
Concern that the outcome, its definition, timing or determination do not match the review question	CONCERN: (low/ high/ unclear)	Low	
<i>Rationale of applicability rating: It was well defined the outcome and it match with the review question.</i>			
DOMAIN 4: Analysis			
Risk of Bias			
Describe numbers of participants, number of candidate predictors (for DEV only), outcome events and events per candidate predictor (for DEV only): A total of 1,475 admissions (by 1,098 unique patients) from 2010 to 2012 were included in this analysis. This group was divided into a derivation cohort (n 1/4 888) and a validation cohort (n 1/4 587). The derivation cohort and validation cohort had similar characteristics. The 30-day all-cause mortality rate was 1.5%. The 30-day HF readmission rate was 9.9%. No information about			

<p><i>Describe how the model was developed (predictor selection, optimism, risk groups, model performance):</i> On multivariable logistic regression analysis, significant independent predictors of 30-day HF readmission were number of previous HF-related admission in the preceding 1 year, index admission length of stay, serum creatinine level, ECG QRS duration, NT-ProBNP level, number of Medical Social Service needs, and b blocker prescription on discharge. Risk for 30-day HF readmission was stratified into low risk, medium risk, and high risk based on predicted readmission rates of <10%, 10% to <30%, and !30%, respectively. The actual observed 30-day HF readmission rates for low risk, medium risk, and high risk were 4%, 15.2%, and 39.5%, respectively. The distribution of the low-, medium-, and high-risk patients in the overall cohort was 55.4%, 37.85%, and 6.8%, respectively.</p>			
<p><i>Describe whether and how the model was validated, either internally (e.g. bootstrapping, cross validation, random split sample) or externally (e.g. temporal validation, geographical validation, different setting, different type of participants):</i> They develop the score and validate it.</p>			
<p><i>Describe the performance measures of the model, e.g. (re)calibration, discrimination, (re)classification, net benefit:</i> The model applied to the entire cohort performed well. The AUC was 0.76. Sensitivity and specificity were 78.3% and 60.7%, respectively. The PPV and NPV were 18.9% and 96%, respectively.</p>			
<p><i>Describe any participants who were excluded from the analysis:</i> Patients were excluded if they had end-stage renal disease (estimated glomerular filtration rate <15 ml/min/1.73 m² or were on dialysis treatment).</p>			
<p><i>Describe missing data on predictors and outcomes as well as methods used for missing data: Do not present any missing data</i></p>			
		Dev	Val
4.1 Were there a reasonable number of participants with the outcome?		Y	
4.2 Were continuous and categorical predictors handled appropriately?		Y	
4.3 Were all enrolled participants included in the analysis?		Y	
4.4 Were participants with missing data handled appropriately?		NI	
4.5 Was selection of predictors based on univariable analysis avoided?		Y	
4.6 Were complexities in the data (e.g. censoring, competing risks, sampling of controls) accounted for appropriately?		NI	
4.7 Were relevant model performance measures evaluated appropriately?		Y	
4.8 Were model overfitting and optimism in model performance accounted for?		Y	
4.9 Do predictors and their assigned weights in the final model correspond to the results from the reported multivariable analysis?		Y	
Risk of bias introduced by the analysis	RISK: (low/ high/ unclear)	Low	
<p><i>Rationale of bias rating: Describe how they get the score</i></p>			

Step 4: Overall assessment

Use the following tables to reach overall judgments about risk of bias and concerns for applicability of the prediction model evaluation (development and/or validation) across all assessed domains.

Reaching an overall judgment about risk of bias of the prediction model evaluation

Low risk of bias	If all domains were rated low risk of bias. If a prediction model was developed without any external validation, and it was rated as <u>low risk of bias for all domains</u> , consider downgrading to high risk of bias . Such a model can only be considered as low risk of bias, if the development was based on a very large data set and included some form of internal validation.
High risk of bias	If at least one domain is judged to be at high risk of bias .
Unclear risk of bias	If an unclear risk of bias was noted in at least one domain and it was low risk for all other domains.

Complete for each evaluation of a distinct model.

Reaching an overall judgment about applicability of the prediction model evaluation

Low concerns for applicability	If low concerns for applicability for all domains, the prediction model evaluation is judged to have low concerns for applicability .
High concerns for applicability	If high concerns for applicability for at least one domain, the prediction model evaluation is judged to have high concerns for applicability .
Unclear concerns for applicability	If unclear concerns (but no “high concern”) for applicability for at least one domain, the prediction model evaluation is judged to have unclear concerns for applicability overall.

Overall judgment about risk of bias and applicability of the prediction model evaluation

Overall judgment of risk of bias	RISK: (low/ high/ unclear)	Low
<i>Summary of sources of potential bias: Low</i>		
Overall judgment of applicability	CONCERN: (low/ high/ unclear)	Low
<i>Summary of applicability concerns: Low</i>		

PROBAST = Prediction model Risk Of Bias ASsessment Tool.

22. Lim NK, et al:

Step 2: Classify the type of prediction model evaluation

Classify the evaluation based on its aim			
Type of prediction study	PROBAST boxes to complete	Tick as appropriate	Definitions for type of prediction model study
Development only	Dev	X	Prediction model development without external validation. These studies may include internal validation methods such as bootstrapping and cross-validation techniques
Development and validation	Dev and Val		Prediction model development combined with external validation in other participants in the same article
Validation only	Val		External validation of existing (previously developed) model in other participants
<i>This table should be completed once for each publication being assessed and for each relevant outcome in your review.</i>			
Publication reference	Lim N-K, et al. Risk prediction for 30-day heart failure-specific readmission or death after discharge: Data from the Korean Acute Heart Failure (KorAHF) registry. Journal of Cardiology (2018). https://doi.org/10.1016/j.jicc.2018.07.009		
Models of interest	risk of 30-day HF-specific readmission or death in HF patients following hospital discharge.		
Outcome of interest	To provide a simple and valid risk score to estimate the 30-day HF-specific readmission or death risk after hospital discharge based on the Korean national HF registry.		

Step 3: Assess risk of bias and applicability

DOMAIN 1: Participants			
A. Risk of Bias			
<i>Describe the sources of data and criteria for participant selection: The Korean Acute Heart Failure registry (KorAHF), supported by the Korea National Institute of Health (KNIH), was established in March 2011. It is an ongoing, prospective, observational study conducted at 10 tertiary hospitals to collect data on patients with acute decompensated heart failure. Information on patient characteristics, treatments, and short- and long-term major outcomes was obtained, and the ethnic and regional profiles of the sample were compared with those of other representative HF registries (ADHERE, ATTEND, OPTIMIZE-HF, EHFSII, THESUS-HF, and ADHERE International).</i>			
		Dev	Val
1.1 Were appropriate data sources used, e.g. cohort, RCT or nested case-control study data?		NI	
1.2 Were all inclusions and exclusions of participants appropriate?		NI	
Risk of bias introduced by selection of participants	RISK: (low/unclear) high/	High	
<i>Rationale of bias rating:</i>			
No inclusion or exclusion criteria were commented.			
B. Applicability			

Describe included participants, setting and dates: Of the 5625 patients, 5341 were aged ± 40 years and included in the study; 775 patients who died during hospitalization or who had missing covariates were excluded. The mean ages of the men and women were 70.3 (± 12.1) years and 72.7 (± 11.3) years, respectively. The observed 30-day HF-specific readmission or death rate after discharge was 9.8% (9.1% in men and 10.5% in women). Patients with a risk of 30-day HF-specific readmission or death were more likely to be older, have a lower body weight, and have a history of HF admission, dilated cardiomyopathy, arrhythmia, cardiac resynchronization therapy, an implantable cardiac defibrillator, chronic obstructive pulmonary disease (COPD), chronic kidney disease, or cancer. Values for LVEF and blood pressure were significantly lower in patients who were readmitted or died. Increased LOS and heavy drinking were also significantly associated with 30-day HF-specific readmission or death. The use of β -blockers, angiotensin-converting enzyme inhibitors or angiotensin II receptor blockers (ACEI/ARB), and diuretics at discharge was lower in patients with a risk of 30-day HF-specific readmission or death.

Concern that the included participants and setting do not match the review question

CONCERN:

(low/ high/
unclear)

High

Rationale of applicability rating: Information from patients were enrolled in at admission from 2011 to 2014.

DOMAIN 2: Predictors

A. Risk of Bias

List and describe predictors included in the final model, e.g. definition and timing of assessment: age, NYHA functional class, clinical history (hypertension, previous HF hospitalization, and COPD), etiology of cardiomyopathy, systolic blood pressure, Na, BNP or NT-proBNP, and prescription of β -blockers and ACEI/ARB at discharge. Making a complex model who resume risk factors and develop this score.

	Dev	Val
2.1 Were predictors defined and assessed in a similar way for all participants?	Y	
2.2 Were predictor assessments made without knowledge of outcome data?	Y	
2.3 Are all predictors available at the time the model is intended to be used?	Y	
Risk of bias introduced by predictors or their assessment	RISK: (low/ high/ unclear)	Low

Rationale of bias rating: The article has a good selection of variables combining variables known as risk factors in ADHF.

B. Applicability

Concern that the definition, assessment or timing of predictors in the model do not match the review question

CONCERN:

(low/ high/
unclear)

Low

Rationale of applicability rating: They use a simple group of variables, easy to get for develop one usefull and very applicable model.

DOMAIN 3: Outcome

A. Risk of Bias

Describe the outcome, how it was defined and determined, and the time interval between predictor assessment and outcome determination: the objective of this study was to provide a simple and valid risk score to estimate the 30-day HF-specific readmission or death risk after hospital discharge based on the Korean national HF registry. Identifying patients with acute heart failure (HF) at high risk for readmission or death after hospital discharge will enable the optimization of treatment and management.

	Dev	Val
3.1 Was the outcome determined appropriately?	Y	

3.2 Was a pre-specified or standard outcome definition used?	Y	
3.3 Were predictors excluded from the outcome definition?	Y	
3.4 Was the outcome defined and determined in a similar way for all participants?	Y	
3.5 Was the outcome determined without knowledge of predictor information?	Y	
3.6 Was the time interval between predictor assessment and outcome determination appropriate?	Y	
Risk of bias introduced by the outcome or its determination	RISK: (low/ high/ unclear)	Low
<i>Rationale of bias rating: It was very well defined outcome, with good model</i>		
B. Applicability		
<i>At what time point was the outcome determined: at one time in the future</i>		
<i>If a composite outcome was used, describe the relative frequency/distribution of each contributing outcome:</i>		
Concern that the outcome, its definition, timing or determination do not match the review question	CONCERN: (low/ high/ unclear)	Low
<i>Rationale of applicability rating: It was demonstrated that this formula is the only Korean population existing. Easy to apply</i>		
DOMAIN 4: Analysis		
Risk of Bias		
<i>Describe numbers of participants, number of candidate predictors (for DEV only), outcome events and events per candidate predictor (for DEV only): 5341 patients. They use 12 predictors. There is no evidence of EPV.</i>		
<i>Describe how the model was developed (predictor selection, optimism, risk groups, model performance): Three regression models used to predict the risk of 30-day HF- specific readmission or death. The complex model, which used multiple logistic regression, included the following: age, NYHA functional class, clinical history (hypertension, previous HF hospitalization, and COPD), etiology of cardiomyopathy, systolic blood pressure, Na, BNP or NT-proBNP, and prescription of b-blockers and ACEI/ARB at discharge. They studied those who were alive after AHF, and developed this score with 30-day readmission or dead.</i>		
<i>Describe whether and how the model was validated, either internally (e.g. bootstrapping, cross validation, random split sample) or externally (e.g. temporal validation, geographical validation, different setting, different type of participants): They develop the score by internal validation.</i>		
<i>Describe the performance measures of the model, e.g. (re)calibration, discrimination, (re)classification, net benefit: The AUCs of the basic, clinical, and complex prediction models based on the logistic regression analyses were 0.647, 0.679, and 0.711, respectively. The C-statistics were significantly increased after the addition of the blood biomarkers ($D = 0.032$, $p < 0.0001$). Additionally, the discrimination ability of the point risk score was almost identical to that for the complex model ($AUC = 0.710$; vs. complex model, $p = 0.3998$) Calibration analysis showed a good level of agreement between the observed risk and the risk predicted by the complex model and point risk score ($p = 0.8019$ and 0.3826, respectively) but not by the basic and clinical models (Fig. 2). The user-category and category-free NRI were 0.17 (95% CI: 0.11– 0.24) and 0.42 (0.33– 0.51), respectively, after the addition of two blood biomarkers to the clinical model</i>		

Describe any participants who were excluded from the analysis: They do not describe excluded patients.			
Describe missing data on predictors and outcomes as well as methods used for missing data: Do not present any missing data			
		Dev	Val
4.1 Were there a reasonable number of participants with the outcome?		Y	
4.2 Were continuous and categorical predictors handled appropriately?		Y	
4.3 Were all enrolled participants included in the analysis?		Y	
4.4 Were participants with missing data handled appropriately?		NI	
4.5 Was selection of predictors based on univariable analysis avoided?		Y	
4.6 Were complexities in the data (e.g. censoring, competing risks, sampling of controls) accounted for appropriately?		NI	
4.7 Were relevant model performance measures evaluated appropriately?		Y	
4.8 Were model overfitting and optimism in model performance accounted for?		Y	
4.9 Do predictors and their assigned weights in the final model correspond to the results from the reported multivariable analysis?		Y	
Risk of bias introduced by the analysis		RISK: (low/ high/ unclear)	Low
Rationale of bias rating: Describe how they get the score			

Step 4: Overall assessment

Use the following tables to reach overall judgments about risk of bias and concerns for applicability of the prediction model evaluation (development and/or validation) across all assessed domains.

Reaching an overall judgment about risk of bias of the prediction model evaluation	
Low risk of bias	If all domains were rated low risk of bias. If a <u>prediction model was developed without any external validation</u> , and it was rated as <u>low risk of bias for all domains</u> , consider downgrading to high risk of bias . Such a model can only be considered as low risk of bias, if the development was based on a very large data set and included some form of internal validation.
High risk of bias	If at least one domain is judged to be at high risk of bias .
Unclear risk of bias	If an unclear risk of bias was noted in at least one domain and it was low risk for all other domains.

Complete for each evaluation of a distinct model.

Reaching an overall judgment about applicability of the prediction model evaluation	
Low concerns for applicability	If low concerns for applicability for all domains, the prediction model evaluation is judged to have low concerns for applicability .
High concerns for applicability	If high concerns for applicability for at least one domain, the prediction model evaluation is judged to have high concerns for applicability .

Unclear concerns for applicability	If unclear concerns (but no “high concern”) for applicability for at least one domain, the prediction model evaluation is judged to have unclear concerns for applicability overall.		
Overall judgment about risk of bias and applicability of the prediction model evaluation			
Overall judgment of risk of bias	RISK: (low/ high/ unclear)		High
Summary of sources of potential bias: High			
Overall judgment of applicability	CONCERN: (low/ high/ unclear)		Low
Summary of applicability concerns: Low			

PROBAST = Prediction model Risk Of Bias ASsessment Tool.

23. Lourenço P, et al:

Step 2: Classify the type of prediction model evaluation

Classify the evaluation based on its aim			
Type of prediction study	PROBAST boxes to complete	Tick as appropriate	Definitions for type of prediction model study
Development only	Dev	X	Prediction model development without external validation. These studies may include internal validation methods such as bootstrapping and cross-validation techniques
Development and validation	Dev and Val		Prediction model development combined with external validation in other participants in the same article
Validation only	Val		External validation of existing (previously developed) model in other participants
<i>This table should be completed once for each publication being assessed and for each relevant outcome in your review.</i>			
Publication reference	Lourenço P, Ribeiro A, Pintalhão M, Silva S, Bettencourt P. Predictors of Six-Month Mortality in BNP-Matched Acute Heart Failure Patients. Am J Cardiol 2015;116:744-748		
Models of interest	Predictors of Six-Month Mortality in BNP-Matched Acute Heart Failure Patients		
Outcome of interest	The aim of this study was to assess prognostic determinants of death in acute HF in B-type natriuretic peptide (BNP)-matched patients.		

Step 3: Assess risk of bias and applicability

DOMAIN 1: Participants			
A. Risk of Bias			
<p><i>Describe the sources of data and criteria for participant selection: We conducted a case-control study to assess mortality predictors in patients with acute HF. To assess prognostic predictors' independence of admission natriuretic peptides, patients surviving and not surviving the 6-month period after discharge were matched for admission BNP. During a 2-year period (January 2009 to December 2010), a registry of acute HF was conducted in the internal medicine department of Hospital São João, Porto, Portugal. All patients admitted with the primary diagnosis of acute HF were eligible for inclusion in the registry. The 2008 European Society of Cardiology guidelines were used for the diagnosis of HF.¹⁵ As part of the registry's protocol, a complete physical examination at admission and discharge was performed; all patients were drawn a venous blood sample within the first 48 hours of hospitalization and in the discharge day; a 12-lead electrocardiogram was performed at admission; and an echocardiogram was performed during hospitalization. Demographic characteristics, medications in use before hospitalization, discharge medication, and co-morbidities were recorded. Plasma BNP was measured with a chemiluminescent immunoassays using an Architect i2000 automated analyzer (Abbott, Lisboa, Portugal). Serum sodium, creatinine, urea, uric acid, albumin, total cholesterol, cardiac troponin I, and C-reactive protein were measured using conventional methods with an Olympus AU5400 automated clinical chemistry analyzer (Beckman- Coulter, Izasa, Porto, Portugal). Prealbumin was assayed using particle-enhanced immunonephelometric assays on a BN II laser nephelometer (Siemens, Lisboa, Portugal). Hemoglobin was obtained using an automated blood counter Sysmex XE- 5000 (Emilio de Azevedo Campos, Porto, Portugal). Heart images were obtained with a standard ultrasound machine (System 6; GE Vingmed, Horten, Norway) with a 2.5-MHz probe. The registry's protocol conforms to the ethical guidelines of the Declaration of Helsinki, and it was approved by the local ethics committee. Patients provided informed consent. Physicians treating patients with acute HF were aware of the ongoing HF registry. Patients' vital status at 6 months was ascertained by consulting hospital registries and by telephone contact with the patients or their relatives.</i></p>			
		De v	Val
1.1 Were appropriate data sources used, e.g. cohort, RCT or nested case-control study data?		Y	
1.2 Were all inclusions and exclusions of participants appropriate?		Y	
Risk of bias introduced by selection of participants	RISK: (low/ high/ unclear)	Low	
<i>Rationale of bias rating: Applied correct criteria for inclusion of participants</i>			
B. Applicability			
<p><i>Describe included participants, setting and dates: A total of 659 patients were included in the registry; from these, 627 survived the hospitalization, and 503 were still alive at 6 months. From this prospectively recruited population, we retrospectively selected a convenience sample of 224 patients (34% of the whole patient population). Cases were 112 patients not surviving the first 6 months after discharge; and controls were patients surviving the follow up period. Controls and cases were 1:1 pair matched for age (± 5 years), gender, and admission BNP.</i></p>			
Concern that the included participants and setting do not match the review question	CONCERN: (low/ high/ unclear)	Low	
<i>Rationale of applicability rating: The characteristics of the study population are consistent with the criteria of the research question.</i>			

DOMAIN 2: Predictors	
A. Risk of Bias	
<p><i>List and describe predictors included in the final model, e.g. definition and timing of assessment: A model was built considering all variables associated with mortality in a univariate approach: atrial fibrillation, arterial hypertension history, admission heart rate and systolic blood pressure, hemoglobin, urea, albumin, and prognostic-modifying therapy.</i></p>	

A second model was built also taking into consideration variables classically associated with outcome: systolic dysfunction, admission New York Heart Association class, and ischemic etiology.			
		Dev	Val
2.1 Were predictors defined and assessed in a similar way for all participants?		Y	
2.2 Were predictor assessments made without knowledge of outcome data?		Y	
2.3 Are all predictors available at the time the model is intended to be used?		Y	
Risk of bias introduced by predictors or their assessment	RISK: (low/ high/ unclear)	Low	
Rationale of bias rating: They adequately describe the way of choosing final predictors.			
B. Applicability			
Concern that the definition, assessment or timing of predictors in the model do not match the review question	CONCERN: (low/ high/ unclear)	High	
Rationale of applicability rating: The evaluation time of the predictive variables and the applicability of these are consistent with the research question of the recent systematic review.			
DOMAIN 3: Outcome			
A. Risk of Bias			
Describe the outcome, how it was defined and determined, and the time interval between predictor assessment and outcome determination: The aim of this study was to assess prognostic determinants of death in acute HF in B-type natriuretic peptide (BNP)-matched patients. 6-month all-cause mortality			
		Dev	Val
3.1 Was the outcome determined appropriately?		Y	
3.2 Was a pre-specified or standard outcome definition used?		Y	
3.3 Were predictors excluded from the outcome definition?		Y	
3.4 Was the outcome defined and determined in a similar way for all participants?		Y	
3.5 Was the outcome determined without knowledge of predictor information?		Y	
3.6 Was the time interval between predictor assessment and outcome determination appropriate?		Y	
Risk of bias introduced by the outcome or its determination	RISK: (low/ high/ unclear)	Low	
Rationale of bias rating: They show an outcome that matches the question of the systematic review..			
B. Applicability			
At what time point was the outcome determined: It was determined during the follow-up in the future.			
If a composite outcome was used, describe the relative frequency/distribution of each contributing outcome:			
Concern that the outcome, its definition, timing or determination do not match the review question	CONCERN: (low/ high/ unclear)	Low	Low

Rationale of applicability rating: The time in which the outcome is determined is consistent with the proposed research question.

DOMAIN 4: Analysis

Risk of Bias

Describe numbers of participants, number of candidate predictors (for DEV only), outcome events and events per candidate predictor (for DEV only): A total of 659 patients were included in the registry; from these, 627 survived the hospitalization, and 503 were still alive at 6 months. From this prospectively recruited population, we retrospectively selected a convenience sample of 224 patients (34% of the whole patient population). Cases were 112 patients not surviving the first 6 months after discharge; and controls were patients surviving the follow up period. Controls and cases were 1:1 pair matched for age (± 5 years), gender, and admission BNP. Candidate variables: 29. EPV: 3.86

Describe how the model was developed (predictor selection, optimism, risk groups, model performance): We conducted a case-control study to assess mortality predictors in patients with acute HF. To assess prognostic predictors' independence of admission natriuretic peptides, patients surviving and not surviving the 6-month period after discharge were matched for admission BNP. During a 2-year period (January 2009 to December 2010), a registry of acute HF was conducted in the internal medicine department of Hospital São João, Porto, Portugal. All patients admitted with the primary diagnosis of acute HF were eligible for inclusion in the registry. The 2008 European Society of Cardiology guidelines were used for the diagnosis of HF.¹⁵ As part of the registry's protocol, a complete physical examination at admission and discharge was performed; all patients were drawn a venous blood sample within the first 48 hours of hospitalization and in the discharge day; a 12-lead electrocardiogram was performed at admission; and an echocardiogram was performed during hospitalization. Demographic characteristics, medications in use before hospitalization, discharge medication, and co-morbidities were recorded. Plasma BNP was measured with a chemiluminescent immunoassays using an Architect i2000 automated analyzer (Abbott, Lisboa, Portugal). Serum sodium, creatinine, urea, uric acid, albumin, total cholesterol, cardiac troponin I, and C-reactive protein were measured using conventional methods with an Olympus AU5400 automated clinical chemistry analyzer (Beckman- Coulter, Izasa, Porto, Portugal). Prealbumin was assayed using particle-enhanced immunonephelometric assays on a BN II laser nephelometer (Siemens, Lisboa, Portugal). Hemoglobin was obtained using an automated blood counter Sysmex XE- 5000 (Emilio de Azevedo Campos, Porto, Portugal). Heart images were obtained with a standard ultrasound machine (System 6; GE Vingmed, Horten, Norway) with a 2.5-MHz probe. The registry's protocol conforms to the ethical guidelines of the Declaration of Helsinki, and it was approved by the local ethics committee. Patients provided informed consent. Physicians treating patients with acute HF were aware of the ongoing HF registry. Patients' vital status at 6 months was ascertained by consulting hospital registries and by telephone contact with the patients or their relatives.

Describe whether and how the model was validated, either internally (e.g. bootstrapping, cross validation, random split sample) or externally (e.g. temporal validation, geographical validation, different setting, different type of participants): The validation method used is not mentioned.

Describe the performance measures of the model, e.g. (re)calibration, discrimination, (re)classification, net benefit: A total of 224 patients were analyzed: 112 surviving and 112 age-, gender-, and BNP matched not surviving a 6-month period post-hospitalization because of acute HF. Ninety-two percent of the patients had >85% admission BNP concordance. The median time to death in nonsurvivors was 69 days (interquartile range 30 to 112 days). The 6-month survivors had higher admission heart rate, higher admission systolic blood pressure, higher hemoglobin, lower plasma urea, and more often had >30% decrease in BNP during hospitalization; they were also more often discharged on HF prognostic-modifying therapy. Mortality predictors in a univariate approach were atrial fibrillation, lower admission systolic blood pressure and heart rate, lower hemoglobin, lower albumin, and higher urea at admission. Hypertension history and prognostic modifying therapy use had a protective effect. A >30% decrease in BNP during hospitalization also showed a protective effect. The article shows 2 multivariate models built to study mortality predictors in our population of age-, gender-, and admission BNP-matched cases (6 months post-discharge nonsurvivors) and controls (alive at 6 months post-discharge). In model 1, variables used were all those associated with outcome in a univariate

<p>approach, and in model 2, variables classically associated with HF outcome-systolic dysfunction, New York Heart Association class, and ischemic etiology—also entered the model. In both multivariate models, no inflammatory, nutritional, or hemodynamic parameters predicted death. The only independent mortality predictor was BNP decrease. In model 1, a BNP decrease >30% during admission predicted a 45% decrease in mortality risk (95% confidence interval 15 to 64), and in model 2, the risk decrease was of 43% (95% confidence interval 11 to 63). If models were built without BNP variation, no variable was associated with mortality.</p>		
<p>Describe any participants who were excluded from the analysis: No characteristics of the excluded subjects are mentioned.</p>		
<p>Describe missing data on predictors and outcomes as well as methods used for missing data: The handling of missing data is not mentioned.</p>		
	Dev	Val
4.1 Were there a reasonable number of participants with the outcome?	N	
4.2 Were continuous and categorical predictors handled appropriately?	PY	
4.3 Were all enrolled participants included in the analysis?	N	
4.4 Were participants with missing data handled appropriately?	NI	
4.5 Was selection of predictors based on univariable analysis avoided?	Y	
4.6 Were complexities in the data (e.g. censoring, competing risks, sampling of controls) accounted for appropriately?	NI	
4.7 Were relevant model performance measures evaluated appropriately?	N	
4.8 Were model overfitting and optimism in model performance accounted for?	NI	
4.9 Do predictors and their assigned weights in the final model correspond to the results from the reported multivariable analysis?	Y	
Risk of bias introduced by the analysis	RISK: (low/ high/ unclear)	High
<p>Rationale of bias rating: No evaluation measures of the models are mentioned, such as calibration or discrimination. Nor do they mention the excluded subjects or the applicable method of the models.</p>		

Step 4: Overall assessment

Use the following tables to reach overall judgments about risk of bias and concerns for applicability of the prediction model evaluation (development and/or validation) across all assessed domains.	
Reaching an overall judgment about risk of bias of the prediction model evaluation	
Low risk of bias	<p>If all domains were rated low risk of bias.</p> <p>If a <u>prediction model was developed without any external validation</u>, and it was rated as <u>low risk of bias for all domains</u>, consider downgrading to high risk of bias. Such a model can only be considered as low risk of bias, if the development was based on a very large data set and included some form of internal validation.</p>
High risk of bias	If at least one domain is judged to be at high risk of bias .
Unclear risk of bias	If an unclear risk of bias was noted in at least one domain and it was low risk for all other domains.
Complete for each evaluation of a distinct model.	
Reaching an overall judgment about applicability of the prediction model evaluation	

Low concerns for applicability	If low concerns for applicability for all domains, the prediction model evaluation is judged to have low concerns for applicability .
High concerns for applicability	If high concerns for applicability for at least one domain, the prediction model evaluation is judged to have high concerns for applicability .
Unclear concerns for applicability	If unclear concerns (but no “high concern”) for applicability for at least one domain, the prediction model evaluation is judged to have unclear concerns for applicability overall.

Overall judgment about risk of bias and applicability of the prediction model evaluation		
Overall judgment of risk of bias	RISK: (low/ high/ unclear)	
Summary of sources of potential bias: High		
Overall judgment of applicability	CONCERN: (low/ high/ unclear)	
Summary of applicability concerns: High		

PROBAST = Prediction model Risk Of Bias ASsessment Tool.

24. Martín-Sánchez J, et al:

Step 2: Classify the type of prediction model evaluation

Classify the evaluation based on its aim			
Type of prediction study	PROBAST boxes to complete	Tick as appropriate	Definitions for type of prediction model study
Development only	Dev	X	Prediction model development without external validation. These studies may include internal validation methods such as bootstrapping and cross-validation techniques
Development and validation	Dev and Val		Prediction model development combined with external validation in other participants in the same article
Validation only	Val		External validation of existing (previously developed) model in other participants

<i>This table should be completed once for each publication being assessed and for each relevant outcome in your review.</i>	
Publication reference	Martín-Sánchez J. Gil V. Llorens P. Herrero P. Jacob J. Fernandez C. et al. Barthel Index—Enhanced Feedback for Effective Cardiac Treatment (BI—EFFECT) Study: Contribution of the Barthel Index to the Heart Failure Risk Scoring System Model in Elderly Adults with Acute Heart Failure in the Emergency Department. The American Geriatrics Society. MARCH 2012—VOL. 60, NO. 3. DOI: 10.1111/j.1532-5415.2011.03845.x
Models of interest	BI-EFFECT
Outcome of interest	Investigate whether the EFFECT scale is applicable to elderly adults with AHF in the ED and whether functional dependence, assessed using the Barthel index (BI), provides additional power to this scale in predicting 30-day mortality.

Step 3: Assess risk of bias and applicability

DOMAIN 1: Participants			
A. Risk of Bias			
Describe the sources of data and criteria for participant selection: The Epidemiology of Acute Heart Failure in the Emergency Department (EAHFE) registry consists of consecutive individuals with AHF diagnosed at 20 Spanish EDs. The participating hospitals are representative of all areas of the country and are tertiary-care university hospitals.			
		Dev	Val
1.1 Were appropriate data sources used, e.g. cohort, RCT or nested case-control study data?		Y	
1.2 Were all inclusions and exclusions of participants appropriate?		Y	
Risk of bias introduced by selection of participants	RISK: (low/ high/ unclear)	Low	
<i>Rationale of bias rating:</i>			
Inclusion criteria was correctly mentioned. No information clearly defined for exclusion criteria			
B. Applicability			
Describe included participants, setting and dates: Of the 2,431 patients included in EAHFE database, 772 were not followed because the hospital was not participating in the short-term outcome study at the time of patient inclusion (n = 690) or the patient refused to authorize follow up (n = 82), 139 were younger than 65, the BI was not available in 47, and essential data were lacking to calculate the EFFECT score in 407, leaving 1,068 patients (44% of the EAHFE registry) for inclusion. Patients included in the BI-EFFECT study differed significantly only in age (>65) and in NYHA class (30% in class III or IV vs 17%; P < .001) from the remaining patients in the EAHFE registry who were not included in the present analysis.			
Concern that the included participants and setting do not match the review question	CONCERN: (low/ high/ unclear)	Low	
<i>Rationale of applicability rating: Information from patients were correctly mentioned.</i>			
DOMAIN 2: Predictors			
A. Risk of Bias			
List and describe predictors included in the final model, e.g. definition and timing of assessment: 17 corresponding to baseline characteristics (age; sex; past medical history of heart failure; systemic hypertension; diabetes mellitus; dyslipidemia; ischemic heart disease; atrial fibrillation; chronic renal failure; cerebrovascular disease; peripheral vasculopathy; chronic obstructive pulmonary disease; dementia; cancer; New York Heart Association (NYHA) class; 19 chronic treatment with beta-blockers, angiotensin converting enzyme inhibitors (ACEIs), or angiotensin receptor antagonists (ARAs); and baseline functional status measured using the BI20) and 11 corresponding to the current AHF episode (cardiac rate, respiratory rate,			

systolic blood pressure, oxygen saturation measured using pulse oximetry, NYHA class for the episode, hemoglobin, sodium, potassium, blood urea nitrogen, troponin, and final destination of the individual) were included.			
		Dev	Val
2.1 Were predictors defined and assessed in a similar way for all participants?		Y	
2.2 Were predictor assessments made without knowledge of outcome data?		Y	
2.3 Are all predictors available at the time the model is intended to be used?		Y	
Risk of bias introduced by predictors or their assessment	RISK: (low/unclear) high/	Low	
<i>Rationale of bias rating: The article has a good selection of variables combining variables known as risk factors in AHF.</i>			
B. Applicability			
Concern that the definition, assessment or timing of predictors in the model do not match the review question	CONCERN: (low/unclear) high/	Low	
<i>Rationale of applicability rating: They use a simple group of variables, easy to get for develop one usefull and very applicable model.</i>			
DOMAIN 3: Outcome			
A. Risk of Bias			
Describe the outcome, how it was defined and determined, and the time interval between predictor assessment and outcome determination: Investigate whether the EFFECT scale is applicable to elderly adults with AHF in the ED and whether functional dependence, assessed using the Barthel index (BI), provides additional power to this scale in predicting 30-day mortality.			
		Dev	Val
3.1 Was the outcome determined appropriately?		Y	
3.2 Was a pre-specified or standard outcome definition used?		Y	
3.3 Were predictors excluded from the outcome definition?		Y	
3.4 Was the outcome defined and determined in a similar way for all participants?		Y	
3.5 Was the outcome determined without knowledge of predictor information?		Y	
3.6 Was the time interval between predictor assessment and outcome determination appropriate?		Y	
Risk of bias introduced by the outcome or its determination	RISK: (low/unclear) high/	Low	
<i>Rationale of bias rating: It was very well defined outcome, with good model</i>			
B. Applicability			
<i>At what time point was the outcome determined: at one time in the future</i>			
<i>If a composite outcome was used, describe the relative frequency/distribution of each contributing outcome:</i>			

Concern that the outcome, its definition, timing or determination do not match the review question	CONCERN: (low/ high/ unclear)	Low	
<i>Rationale of applicability rating: It was well defined the outcome and it match with the review question.</i>			
DOMAIN 4: Analysis			
Risk of Bias			
<i>Describe numbers of participants, number of candidate predictors (for DEV only), outcome events and events per candidate predictor (for DEV only): 1,068 patients (44% of the EAHFE registry) for inclusion.</i>			
<p><i>Describe how the model was developed (predictor selection, optimism, risk groups, model performance):</i> BI = 61–90), and severe dependence (BI 60 points). Risk of 30-day mortality based on EFFECT score was classified into five stages: very low (60 points), low (61–90 points), intermediate (91–120 points), high (121–150 points), and very high (>150 points). For some analyses, dependence was grouped as severe (BI 60 points) or nonsevere (BI > 60 points) and the risk groups of the EFFECT scale as very low or low (90 points), intermediate (91–120 points), or high or very high (>120 points).</p> <p>Thirty-day mortality was obtained through the clinical history of each patient or with a direct telephone call to the patient or a relative 31 to 60 days after discharge.</p>			
<i>Describe whether and how the model was validated, either internally (e.g. bootstrapping, cross validation, random split sample) or externally (e.g. temporal validation, geographical validation, different setting, different type of participants): They develop the score with the EAHFE registry in 20 spanish ED.</i>			
<i>Describe the performance measures of the model, e.g. (re)calibration, discrimination, (re)classification, net benefit: It found that the EFFECT scale has good discriminatory capacity, with an AUC of 0.69 (95% CI = 0.62–0.76).</i>			
<i>Describe any participants who were excluded from the analysis: No mention of the exclusion criteria..</i>			
<i>Describe missing data on predictors and outcomes as well as methods used for missing data: Do not present missing data procedure. Only 3% of data related to the 28 independent variables were missing</i>			
		Dev	Val
4.1 Were there a reasonable number of participants with the outcome?		Y	
4.2 Were continuous and categorical predictors handled appropriately?		Y	
4.3 Were all enrolled participants included in the analysis?		N	
4.4 Were participants with missing data handled appropriately?		N	
4.5 Was selection of predictors based on univariable analysis avoided?		Y	
4.6 Were complexities in the data (e.g. censoring, competing risks, sampling of controls) accounted for appropriately?		NI	
4.7 Were relevant model performance measures evaluated appropriately?		Y	
4.8 Were model overfitting and optimism in model performance accounted for?		Y	
4.9 Do predictors and their assigned weights in the final model correspond to the results from the reported multivariable analysis?		Y	

Risk of bias introduced by the analysis	RISK: (low/ high/ unclear)	Unclear	
<i>Rationale of bias rating: Describe how they get the score</i>			

Step 4: Overall assessment

Use the following tables to reach overall judgments about risk of bias and concerns for applicability of the prediction model evaluation (development and/or validation) across all assessed domains.

Reaching an overall judgment about risk of bias of the prediction model evaluation	
Low risk of bias	If all domains were rated low risk of bias. If a prediction model was developed without any external validation, and it was rated as <u>low risk of bias for all domains</u> , consider downgrading to high risk of bias . Such a model can only be considered as low risk of bias, if the development was based on a very large data set and included some form of internal validation.
High risk of bias	If at least one domain is judged to be at high risk of bias .
Unclear risk of bias	If an unclear risk of bias was noted in at least one domain and it was low risk for all other domains.

Complete for each evaluation of a distinct model.

Reaching an overall judgment about applicability of the prediction model evaluation	
Low concerns for applicability	If low concerns for applicability for all domains, the prediction model evaluation is judged to have low concerns for applicability .
High concerns for applicability	If high concerns for applicability for at least one domain, the prediction model evaluation is judged to have high concerns for applicability .
Unclear concerns for applicability	If unclear concerns (but no “high concern”) for applicability for at least one domain, the prediction model evaluation is judged to have unclear concerns for applicability overall.

Overall judgment about risk of bias and applicability of the prediction model evaluation		
Overall judgment of risk of bias	RISK: (low/ high/ unclear)	Unclear
<i>Summary of sources of potential bias: Unclear</i>		
Overall judgment of applicability	CONCERN: (low/ high/ unclear)	Low
<i>Summary of applicability concerns: Low</i>		

PROBAST = Prediction model Risk Of Bias ASsessment Tool.

25. Miro O, et al:

Step 2: Classify the type of prediction model evaluation

Classify the evaluation based on its aim			
Type of prediction study	PROBAST boxes to complete	Tick as appropriate	Definitions for type of prediction model study
Development only	Dev		Prediction model development without external validation. These studies may include internal validation methods such as bootstrapping and cross-validation techniques
Development and validation	Dev and Val	X	Prediction model development combined with external validation in other participants in the same article
Validation only	Val		External validation of existing (previously developed) model in other participants
<i>This table should be completed once for each publication being assessed and for each relevant outcome in your review.</i>			
Publication reference	Miro, O; Rossello, X; Gil, V; Martin-Sanchez, FJ; Llorens, P; Herrero-Puente, P; Jacob, J; Bueno, H; Pocock, SJ; Grp, I.-, SR (2017) Predicting 30-Day Mortality for Patients With Acute Heart Failure in the Emergency Department A Cohort Study. Annals of internal medicine DOI: https://doi.org/10.7326/M16-2726		
Models of interest	MEESSI-AHF		
Outcome of interest	To predict mortality using data that are readily available at ED admission.		

Step 3: Assess risk of bias and applicability

DOMAIN 1: Participants			
A. Risk of Bias			
Describe the sources of data and criteria for participant selection: The EAHFE Registry collects detailed information on consecutive patients attending 34 Spanish EDs with a final diagnosis of AHF. Hospitals participate in the EAHFE Registry voluntarily, and they include university and community hospitals, EDs with high, medium or low volume of attendances and hospitals from all areas of the country. Attending emergency physicians use Framingham's clinical diagnostic criteria to identify patients for the registry. Thereafter, the diagnosis is double-checked by the principal investigator of each centre, who makes the final adjudication of AHF diagnosis based on the review of medical charts and all complementary tests done during the ED stay and any hospitalization. The diagnosis was confirmed by natriuretic peptide determinations or echocardiography in the 92% of patients included in the EAHFE Registry.			
		Dev	Val
1.1 Were appropriate data sources used, e.g. cohort, RCT or nested case-control study data?		Y	Y
1.2 Were all inclusions and exclusions of participants appropriate?		Y	Y
Risk of bias introduced by selection of participants	RISK: (low/ high/ unclear)	Low	Low
<i>Rationale of bias rating:</i>			
Inclusion criteria was correctly mentioned. No information clearly defined for exclusion criteria			
B. Applicability			
<i>Describe included participants, setting and dates: The study population consisted of patients with AHF.</i>			
Concern that the included participants and setting do not match the review question	CONCERN: (low/ high/ unclear)	Low	Low
<i>Rationale of applicability rating: Information from patients were correctly mentioned.</i>			
DOMAIN 2: Predictors			
A. Risk of Bias			
<i>List and describe predictors included in the final model, e.g. definition and timing of assessment: Barthel Index, SBP, Age, NT-proBNP, Potassium level, Positive troponin level, NYHA class, Respiratory rate, Low-output symptoms, oxygen saturation, episode associated with ACS, Hypertrophy on ECG, Creatinine level.</i>			
		Dev	Val
2.1 Were predictors defined and assessed in a similar way for all participants?		Y	Y
2.2 Were predictor assessments made without knowledge of outcome data?		Y	Y
2.3 Are all predictors available at the time the model is intended to be used?		Y	Y
Risk of bias introduced by predictors or their assessment	RISK: (low/ high/ unclear)	Low	Low
<i>Rationale of bias rating: The article has a good selection of variables combining variables known as risk factors in AHF.</i>			
B. Applicability			
Concern that the definition, assessment or timing of predictors in the model do not match the review question	CONCERN: (low/ high/ unclear)	Low	Low
<i>Rationale of applicability rating: They use a simple group of variables, easy to get for develop one usefull and very applicable model.</i>			

DOMAIN 3: Outcome			
A. Risk of Bias			
Describe the outcome, how it was defined and determined, and the time interval between predictor assessment and outcome determination: To predict mortality using data that are readily available at ED admission.			
		Dev	Val
3.1 Was the outcome determined appropriately?		Y	Y
3.2 Was a pre-specified or standard outcome definition used?		Y	Y
3.3 Were predictors excluded from the outcome definition?		Y	Y
3.4 Was the outcome defined and determined in a similar way for all participants?		Y	Y
3.5 Was the outcome determined without knowledge of predictor information?		Y	NI
3.6 Was the time interval between predictor assessment and outcome determination appropriate?		Y	Y
Risk of bias introduced by the outcome or its determination	RISK: (low/ high/ unclear)	Low	Low
<i>Rationale of bias rating: It was very well defined outcome, with good model</i>			
B. Applicability			
<i>At what time point was the outcome determined: at one time in the future</i>			
<i>If a composite outcome was used, describe the relative frequency/distribution of each contributing outcome:</i>			
Concern that the outcome, its definition, timing or determination do not match the review question	CONCERN: (low/ high/ unclear)	Low	Low
<i>Rationale of applicability rating: It was well defined the outcome and it match with the review question.</i>			
DOMAIN 4: Analysis			
Risk of Bias			
Describe numbers of participants, number of candidate predictors (for DEV only), outcome events and events per candidate predictor (for DEV only): 4867 patients were included In the model. identified 88 candidate predictor variables. No information about EPV.			
Describe how the model was developed (predictor selection, optimism, risk groups, model performance): We used a logistic regression model with forward stepwise variable selection to identify 13 highly significant independent predictors of death that we then included in the MEESSEI-AHF risk score.			
Describe whether and how the model was validated, either internally (e.g. bootstrapping, cross validation, random split sample) or externally (e.g. temporal validation, geographical validation, different setting, different type of participants): They develop the score and validate it.			
Describe the performance measures of the model, e.g. (re)calibration, discrimination, (re)classification, net benefit: Thirteen independent risk factors were identified in the derivation cohort and were combined into an overall score, the MEESSEI-AHF (Multiple Estimation of risk based on the Emergency department Spanish Score In patients with AHF) score. This score predicted 30-day mortality with excellent discrimination (c-statistic, 0.836) and calibration (Hosmer–Lemeshow P = 0.99) and provided a steep gradient in 30-day			

mortality across risk groups (<2% for patients in the 2 lowest risk quintiles and 45% in the highest risk decile). These characteristics were confirmed in the validation cohort (c-statistic, 0.828). Multiple sensitivity analyses did not find important amounts of confounding or bias.			
Describe any participants who were excluded from the analysis: We excluded patients who had a concurrent diagnosis of ST-segment elevation myocardial infarction (approximately 3%).			
Describe missing data on predictors and outcomes as well as methods used for missing data: Some significant predictors had a high number of missing values, which we have addressed with multiple imputation techniques and sensitivity analyses.			
		Dev	Val
4.1 Were there a reasonable number of participants with the outcome?		Y	Y
4.2 Were continuous and categorical predictors handled appropriately?		Y	Y
4.3 Were all enrolled participants included in the analysis?		Y	Y
4.4 Were participants with missing data handled appropriately?		N	N
4.5 Was selection of predictors based on univariable analysis avoided?		Y	Y
4.6 Were complexities in the data (e.g. censoring, competing risks, sampling of controls) accounted for appropriately?		NI	NI
4.7 Were relevant model performance measures evaluated appropriately?		Y	Y
4.8 Were model overfitting and optimism in model performance accounted for?		Y	Y
4.9 Do predictors and their assigned weights in the final model correspond to the results from the reported multivariable analysis?		Y	Y
Risk of bias introduced by the analysis	RISK: (low/ unclear) high/	Low	Low
Rationale of bias rating: Describe how they get the score			

Step 4: Overall assessment

Use the following tables to reach overall judgments about risk of bias and concerns for applicability of the prediction model evaluation (development and/or validation) across all assessed domains.

Reaching an overall judgment about risk of bias of the prediction model evaluation	
Low risk of bias	If all domains were rated low risk of bias. If a prediction model was developed without any external validation, and it was rated as <u>low risk of bias for all domains</u> , consider downgrading to high risk of bias . Such a model can only be considered as low risk of bias, if the development was based on a very large data set and included some form of internal validation.
High risk of bias	If at least one domain is judged to be at high risk of bias .
Unclear risk of bias	If an unclear risk of bias was noted in at least one domain and it was low risk for all other domains.

Complete for each evaluation of a distinct model.

Reaching an overall judgment about applicability of the prediction model evaluation	
Low concerns for applicability	If low concerns for applicability for all domains, the prediction model evaluation is judged to have low concerns for applicability .
High concerns for applicability	If high concerns for applicability for at least one domain, the prediction model evaluation is judged to have high concerns for applicability .
Unclear concerns for applicability	If unclear concerns (but no “high concern”) for applicability for at least one domain, the prediction model evaluation is judged to have unclear concerns for applicability overall.

Overall judgment about risk of bias and applicability of the prediction model evaluation		
Overall judgment of risk of bias	RISK: (low/ high/ unclear)	Low
<i>Summary of sources of potential bias: Low</i>		
Overall judgment of applicability	CONCERN: (low/ high/ unclear)	Low
<i>Summary of applicability concerns: Low</i>		

PROBAST = Prediction model Risk Of Bias ASsessment Tool.

26. Nakada Y, et al:

Step 2: Classify the type of prediction model evaluation

Classify the evaluation based on its aim			
Type of prediction study	PROBAST boxes to complete	Tick as appropriate	Definitions for type of prediction model study
Development only	Dev		Prediction model development without external validation. These studies may include internal validation methods such as bootstrapping and cross-validation techniques
Development and validation	Dev and Val	X	Prediction model development combined with external validation in other participants in the same article
Validation only	Val		External validation of existing (previously developed) model in other participants
<i>This table should be completed once for each publication being assessed and for each relevant outcome in your review.</i>			
Publication reference	Nakada Y, Kawakami R, Matsushima S, Ide T, Kanaoka K, et al. Simple Risk Score to Predict Survival in Acute Decompensated Heart Failure. Circulation Journal (2018). doi:10.1253/circj.CJ-18-1116.		
Models of interest	A2B Score		
Outcome of interest	Designed a more fundamental and a simpler score system for patients with ADHF.		

Step 3: Assess risk of bias and applicability

DOMAIN 1: Participants		
A. Risk of Bias		
Describe the sources of data and criteria for participant selection: The present study used 2 HF cohorts, the Nara registry and analyses for heart failure (NARA-HF) 3 study and the Japanese Cardiac Registry of Heart Failure in Cardiology (JCARE-CARD).		
	De	Val

		v	
1.1 Were appropriate data sources used, e.g. cohort, RCT or nested case-control study data?		Y	Y
1.2 Were all inclusions and exclusions of participants appropriate?		Y	Y
Risk of bias introduced by selection of participants	RISK: (low/unclear) high/	Low	Low
<i>Rationale of bias rating:</i>			
Inclusion and exclusion criteria were well defined in both groups.			
B. Applicability			
<i>Describe included participants, setting and dates: The NARA-HF 3 study was used as a derivation model, and the JCARE-CARD was used as a validation cohort. The NARA-HF 3 study recruited 1,074 consecutive patients following emergency admission for ADHF between January 2007 and December 2016. The JCARE-CARD have 2,675 enrolled patients, we analyzed 542 patients who had discharge data and 2-year follow-up data. Patients with acute myocardial infarction, acute myocarditis, or acute HF with acute pulmonary were excluded. we excluded 45 who died during the current hospitalization, 24 who were lost to follow-up, 146 who were in the 2-year follow-up period, and 120 with missing discharge data. Thus, we analyzed 739 patients with ADHF.</i>			
Concern that the included participants and setting do not match the review question	CONCERN: (low/unclear) high/	Low	Low
<i>Rationale of applicability rating: Adequated description of patients selection</i>			

DOMAIN 2: Predictors			
A. Risk of Bias			
<i>List and describe predictors included in the final model, e.g. definition and timing of assessment: Age, Male, BMI, SBP, LVEF, eGFR, Hemoglobin and BNP. Age, Hb, and plasma BNP at discharge were found to be strong independent risk factors for the primary outcome in the multivariable Cox proportional hazards model.</i>			
		De v	Val
2.1 Were predictors defined and assessed in a similar way for all participants?		Y	Y
2.2 Were predictor assessments made without knowledge of outcome data?		Y	Y
2.3 Are all predictors available at the time the model is intended to be used?		Y	Y
Risk of bias introduced by predictors or their assessment	RISK: (low/unclear) high/	Low	Low
<i>Rationale of bias rating: The article used those 3 strong variables whose were found in the multivariated model.</i>			
B. Applicability			
Concern that the definition, assessment or timing of predictors in the model do not match the review question	CONCERN: (low/unclear) high/	Low	Low
<i>Rationale of applicability rating: they used 3 simple variables easy to get and to apply.</i>			
DOMAIN 3: Outcome			
A. Risk of Bias			

Describe the outcome, how it was defined and determined, and the time interval between predictor assessment and outcome determination: Designed a more fundamental and a simpler score system for patients with ADHF because previous ones are not well-applied.			
		Dev	Val
3.1 Was the outcome determined appropriately?		Y	Y
3.2 Was a pre-specified or standard outcome definition used?		Y	Y
3.3 Were predictors excluded from the outcome definition?		Y	Y
3.4 Was the outcome defined and determined in a similar way for all participants?		Y	Y
3.5 Was the outcome determined without knowledge of predictor information?		Y	Y
3.6 Was the time interval between predictor assessment and outcome determination appropriate?		NI	NI
Risk of bias introduced by the outcome or its determination	RISK: (low/ high/ unclear)	Low	Low
Rationale of bias rating: It was very well defined outcome.			
B. Applicability			
At what time point was the outcome determined: at one time in the future			
If a composite outcome was used, describe the relative frequency/distribution of each contributing outcome:			
Concern that the outcome, its definition, timing or determination do not match the review question	CONCERN: (low/ high/ unclear)	Low	Low
Rationale of applicability rating: They had 100% follow up population.			

DOMAIN 4: Analysis			
Risk of Bias			
Describe numbers of participants, number of candidate predictors (for DEV only), outcome events and events per candidate predictor (for DEV only): 739 patients in Derivation cohort and 542 patients in Validation Cohort. It has 8 candidate predictors: Age, Male, BMI, SBP, LVEF, eGFR, Hemoglobin and BNP. Age, Hb, and plasma BNP			
Describe how the model was developed (predictor selection, optimism, risk groups, model performance): Age, Hb, and plasma BNP at discharge were found to be strong independent risk factors for the primary outcome in the multivariable Cox proportional hazards model. The risk of the primary outcome was found to increase with older age, lower Hb, and higher plasma BNP.			
Describe whether and how the model was validated, either internally (e.g. bootstrapping, cross validation, random split sample) or externally (e.g. temporal validation, geographical validation, different setting, different type of participants): they developed in the derivation cohort and validated internally in the validation cohort.			
Describe the performance measures of the model, e.g. (re)calibration, discrimination, (re)classification, net benefit: Calibration of the final model was tested using the Grønnesby-Borgan goodness-of-fit statistic. The discrimination was shown with AUC of 0.711 for group 7 (Age, Hb and BNP). The net reclassification was performed by NRI whose value had the highest value against group 3 (BNP alone) of 0.502. No information about net benefit was presented.			
Describe any participants who were excluded from the analysis: acute myocardial infarction, acute myocarditis, or acute HF with acute pulmonary embolism were excluded.			
Describe missing data on predictors and outcomes as well as methods used for missing data: Do not present any missing data because they were excluded previously.			
		Dev	Val
4.1 Were there a reasonable number of participants with the outcome?		Y	Y
4.2 Were continuous and categorical predictors handled appropriately?		Y	Y
4.3 Were all enrolled participants included in the analysis?		Y	Y
4.4 Were participants with missing data handled appropriately?		Y	Y
4.5 Was selection of predictors based on univariable analysis avoided?		Y	Y
4.6 Were complexities in the data (e.g. censoring, competing risks, sampling of controls) accounted for appropriately?		Y	Y
4.7 Were relevant model performance measures evaluated appropriately?		Y	Y
4.8 Were model overfitting and optimism in model performance accounted for?		Y	Y
4.9 Do predictors and their assigned weights in the final model correspond to the results from the reported multivariable analysis?		Y	Y
Risk of bias introduced by the analysis	RISK: (low/ unclear) high/	Low	Low
Rationale of bias rating: Describe correctly the way they chose strong predictors.			

Step 4: Overall assessment

Use the following tables to reach overall judgments about risk of bias and concerns for applicability of the prediction model evaluation (development and/or validation) across all assessed domains.

Reaching an overall judgment about risk of bias of the prediction model evaluation	
Low risk of bias	If all domains were rated low risk of bias. If a prediction model was developed without any external validation, and it was rated as <u>low risk of bias for all domains</u> , consider downgrading to high risk of bias . Such a model can only be considered as low risk of bias, if the development was based on a very large data set and included some form of internal validation.
High risk of bias	If at least one domain is judged to be at high risk of bias .
Unclear risk of bias	If an unclear risk of bias was noted in at least one domain and it was low risk for all other domains.

Complete for each evaluation of a distinct model.

Reaching an overall judgment about applicability of the prediction model evaluation	
Low concerns for applicability	If low concerns for applicability for all domains, the prediction model evaluation is judged to have low concerns for applicability .
High concerns for applicability	If high concerns for applicability for at least one domain, the prediction model evaluation is judged to have high concerns for applicability .
Unclear concerns for applicability	If unclear concerns (but no “high concern”) for applicability for at least one domain, the prediction model evaluation is judged to have unclear concerns for applicability overall.

Overall judgment about risk of bias and applicability of the prediction model evaluation		
Overall judgment of risk of bias	RISK: (low/ high/ unclear)	Low
<i>Summary of sources of potential bias: Low</i>		
Overall judgment of applicability	CONCERN: (low/ high/ unclear)	Low
<i>Summary of applicability concerns: Low</i>		

PROBAST = Prediction model Risk Of Bias ASsessment Tool.

27. O'Connor C, et al:

Step 2: Classify the type of prediction model evaluation

Classify the evaluation based on its aim			
Type of prediction study	PROBAST boxes to complete	Tick as appropriate	Definitions for type of prediction model study
Development only	Dev		Prediction model development without external validation. These studies may include internal validation methods such as bootstrapping and cross-validation techniques
Development and validation	Dev and Val	X	Prediction model development combined with external validation in other participants in the same article
Validation only	Val		External validation of existing (previously developed) model in other participants
<i>This table should be completed once for each publication being assessed and for each relevant outcome in your review.</i>			
Publication reference	O'Connor C, Mentz R, Cotter G, Metra M, Cleland J. The PROTECT in-hospital risk model: 7-day outcome in patients hospitalized with acute heart failure and renal dysfunction. European Journal of Heart Failure (2012) 14, 605–612 doi:10.1093/eurjhf/hfs029		
Models of interest	PROTECT in-hospital risk model		
Outcome of interest	To identify the predictors of 7-day heart failure events or death in patients hospitalized with AHF.		

Step 3: Assess risk of bias and applicability

DOMAIN 1: Participants
A. Risk of Bias
Describe the sources of data and criteria for participant selection: The international PROTECT trial enrolled 2033 patients hospitalized with AHF and mild or moderate impairment of renal function. The inclusion criteria were: (i) history of HF treated for at least 14 days with diuretic therapy; (ii) hospitalization for AHF requiring i.v. diuretic therapy; (iii) admission creatinine clearance 20 – 80 mL/min; and (iv) systolic blood pressure ≥ 95 mmHg

	De v	Val
1.1 Were appropriate data sources used, e.g. cohort, RCT or nested case-control study data?	Y	Y
1.2 Were all inclusions and exclusions of participants appropriate?	Y	Y
Risk of bias introduced by selection of participants	RISK: (low/ high/ unclear)	Low Low
<i>Rationale of bias rating:</i>		
Inclusion and exclusion criteria was correctly mentioned.		
B. Applicability		
<i>Describe included participants, setting and dates: The study population consisted of patients with AHF.</i>		
Concern that the included participants and setting do not match the review question	CONCERN: (low/ high/ unclear)	Low Low
<i>Rationale of applicability rating: Information from patients were correctly mentioned.</i>		

DOMAIN 2: Predictors		
A. Risk of Bias		
<i>List and describe predictors included in the final model, e.g. definition and timing of assessment: BUN, SBP, Respiratory rate, Hospitalization for HF in the past year, albumin, heart rate, cholesterol, diabetes.</i>		
	De v	Val
2.1 Were predictors defined and assessed in a similar way for all participants?	Y	Y
2.2 Were predictor assessments made without knowledge of outcome data?	Y	Y
2.3 Are all predictors available at the time the model is intended to be used?	Y	Y
Risk of bias introduced by predictors or their assessment	RISK: (low/ high/ unclear)	Low Low
<i>Rationale of bias rating: The article has a good selection of variables combining variables known as risk factors in AHF.</i>		
B. Applicability		
Concern that the definition, assessment or timing of predictors in the model do not match the review question	CONCERN: (low/ high/ unclear)	Low Low
<i>Rationale of applicability rating: They use a simple group of variables, easy to get for develop one usefull and very applicable model.</i>		
DOMAIN 3: Outcome		
A. Risk of Bias		
Describe the outcome, how it was defined and determined, and the time interval between predictor assessment and outcome determination: To identify the predictors of 7-day heart failure events or death in patients hospitalized with AHF.		
	Dev	Val
3.1 Was the outcome determined appropriately?	Y	Y
3.2 Was a pre-specified or standard outcome definition used?	Y	Y
3.3 Were predictors excluded from the outcome definition?	Y	Y
3.4 Was the outcome defined and determined in a similar way for all participants?	Y	Y
3.5 Was the outcome determined without knowledge of predictor information?	Y	NI

3.6 Was the time interval between predictor assessment and outcome determination appropriate?	Y	Y
Risk of bias introduced by the outcome or its determination	RISK: (low/unclear) high/	Low Low
<i>Rationale of bias rating: It was very well defined outcome, with good model</i>		
B. Applicability		
<i>At what time point was the outcome determined: at one time in the future</i>		
<i>If a composite outcome was used, describe the relative frequency/distribution of each contributing outcome:</i>		
Concern that the outcome, its definition, timing or determination do not match the review question	CONCERN: (low/unclear) high/	Low Low
<i>Rationale of applicability rating: It was well defined the outcome and it match with the review question.</i>		
DOMAIN 4: Analysis		
Risk of Bias		
<i>Describe numbers of participants, number of candidate predictors (for DEV only), outcome events and events per candidate predictor (for DEV only):</i> 2015 patients. They mentioned the presence of candidate predictors but not available in the article. No information of EPV.		
<i>Describe how the model was developed (predictor selection, optimism, risk groups, model performance):</i> The multivariable model was used to develop a risk score for pre- dicting the composite endpoint at 7 days. A weighted integer was assigned to each independent predictor, based on the predictor's coefficient in the regression model. For each patient, the weighted integers were summed to obtain a total risk score with a range 0–100 points. The discrimination of the risk score was assessed with a c-statistic and the calibration by comparing observed vs. estimated event rates. A histogram of the distribution of the risk score in the study population was generated.		
<i>Describe whether and how the model was validated, either internally (e.g. bootstrapping, cross validation, random split sample) or externally (e.g. temporal validation, geographical validation, different setting, different type of participants):</i> They develop the score and validate it. The model was internally validated using bootstrap re-sampling (c-index 0.67). The VERITAS study population used for external validation was similar to the PROTECT population, with enrolment of 40% women and 8% African American race with a mean age of 70 years, and mean creatinine and LVEF of □ 1.3 mg/dL and 29%, respectively.		
<i>Describe the performance measures of the model, e.g. (re)calibration, discrimination, (re)classification, net benefit:</i> The risk score demonstrated modest discrimination (c-statistic of 0.67). The calibration was good, as demonstrated by the Hosmer – May test (P 1/4 0.67) and the plot of observed vs. estimated event rates by deciles of risk		
<i>Describe any participants who were excluded from the analysis:</i> Key exclusion criteria were: (i) use of inotropic agents or mechanical support; (ii) N-terminal pro brain natriuretic peptide (NT-proBNP) ,2000 pg/mL; (iii) severe pulmonary disease; (iv) signifi- cant stenotic valvular disease, recent acute coronary syndrome, or sig- nificant arrhythmias; (v) heart transplant recipient; and (vi) history of seizure/stroke within 2 years.		
<i>Describe missing data on predictors and outcomes as well as methods used for missing data:</i> Most of the variables used in the multivariate modelling had no or ,1% missing values. Only five variables had between 5% and 10% missing values. For the purpose of developing the multivariable model, 25 data sets were generated by multiple imputation of missing values.		
	Dev	Val

4.1 Were there a reasonable number of participants with the outcome?	Y	Y
4.2 Were continuous and categorical predictors handled appropriately?	Y	Y
4.3 Were all enrolled participants included in the analysis?	Y	Y
4.4 Were participants with missing data handled appropriately?	Y	Y
4.5 Was selection of predictors based on univariable analysis avoided?	Y	Y
4.6 Were complexities in the data (e.g. censoring, competing risks, sampling of controls) accounted for appropriately?	NI	NI
4.7 Were relevant model performance measures evaluated appropriately?	Y	Y
4.8 Were model overfitting and optimism in model performance accounted for?	Y	Y
4.9 Do predictors and their assigned weights in the final model correspond to the results from the reported multivariable analysis?	Y	Y
Risk of bias introduced by the analysis	RISK: (low/ high/ unclear)	Low Low
<i>Rationale of bias rating: Describe how they get the score</i>		

Step 4: Overall assessment

Use the following tables to reach overall judgments about risk of bias and concerns for applicability of the prediction model evaluation (development and/or validation) across all assessed domains.

Reaching an overall judgment about risk of bias of the prediction model evaluation	
Low risk of bias	If all domains were rated low risk of bias. If a prediction model was developed without any external validation, and it was rated as <u>low risk of bias for all domains</u> , consider downgrading to high risk of bias . Such a model can only be considered as low risk of bias, if the development was based on a very large data set and included some form of internal validation.
High risk of bias	If at least one domain is judged to be at high risk of bias .
Unclear risk of bias	If an unclear risk of bias was noted in at least one domain and it was low risk for all other domains.

Complete for each evaluation of a distinct model.

Reaching an overall judgment about applicability of the prediction model evaluation	
Low concerns for applicability	If low concerns for applicability for all domains, the prediction model evaluation is judged to have low concerns for applicability .
High concerns for applicability	If high concerns for applicability for at least one domain, the prediction model evaluation is judged to have high concerns for applicability .
Unclear concerns for applicability	If unclear concerns (but no “high concern”) for applicability for at least one domain, the prediction model evaluation is judged to have unclear concerns for applicability overall.

Overall judgment about risk of bias and applicability of the prediction model evaluation		
Overall judgment of risk of bias	RISK: (low/ high/ unclear)	Low

Summary of sources of potential bias: Low		
Overall judgment of applicability	CONCERN: (low/ high/ unclear)	Low
Summary of applicability concerns: Low		

PROBAST = Prediction model Risk Of Bias ASsessment Tool.

28. Okazaki H, et al:

Step 2: Classify the type of prediction model evaluation

Classify the evaluation based on its aim			
Type of prediction study	PROBAST boxes to complete	Tick as appropriate	Definitions for type of prediction model study
Development only	Dev	X	Prediction model development without external validation. These studies may include internal validation methods such as bootstrapping and cross-validation techniques
Development and validation	Dev and Val		Prediction model development combined with external validation in other participants in the same article
Validation only	Val		External validation of existing (previously developed) model in other participants
<i>This table should be completed once for each publication being assessed and for each relevant outcome in your review.</i>			
Publication reference	Okazaki H, Shirakabe A, Hata N, Yamamoto M, Kobayashi N, Shinada T, et al. New scoring system (APACHE-HF) for predicting adverse outcomes in patients with acute heart failure: Evaluation of the APACHEII and Modified APACHEII scoring systems. Journal of Cardiology. 2014; 64: 441–449		
Models of interest	APACHE-HF		
Outcome of interest	Predict adverse outcomes in patients with AHF		

Step 3: Assess risk of bias and applicability

DOMAIN 1: Participants
A. Risk of Bias
<i>Describe the sources of data and criteria for participant selection: Clinical data were collected from 824 patients with AHF who were admitted to the intensive care unit at Chiba Hokusoh Hospital, Nippon Medical School between January 2000 and July 2012. AHF was defined as either new-onset HF or decompensation of chronic HF with symptoms sufficient to warrant hospitalization [19]. HF was diagnosed according to the Framingham criteria for a clinical diagnosis of HF based on the satisfaction of two major criteria or one major and two minor criteria [20]. All patients had a New York Heart Association (NYHA) functional class of either Class III or IV. AHF patients with one of the following criteria were admitted to the intensive care unit(ICU) by physician's</i>

<p>decision in the present study: (1)patients who need high projectile oxygen inhalation (including mechanical support) to treat orthopnea,(2) patients who need intropo or mechanical support with low blood pressure,(3) patients who need many types of diuretics to improve the general or lung edema. Patients with HF caused by acute coronary syndrome were excluded from the study. All data were retrospectively retrieved from hospital medical records.</p>			
		De v	Val
1.1 Were appropriate data sources used, e.g. cohort, RCT or nested case-control study data?		Y	
1.2 Were all inclusions and exclusions of participants appropriate?		Y	
Risk of bias introduced by selection of participants	RISK: (low/ high/ unclear)	Low	
<p><i>Rationale of bias rating: Applied correct criteria for inclusion of participants</i></p>			
B. Applicability			
<p><i>Describe included participants, setting and dates:</i> AHF patients were divided into two groups according to in-hospital mortality : the alive group(n = 750) and the dead group (n = 74). The patient cohort included 67.2% male subjects, with a median age of 74 years. The systolic blood pressure (BP) values were significantly lower, the number of NYHA class IV patients was significantly higher, the serum hemoglobin levels were significantly lower, the serum urinary acid levels were significantly higher, the serum BUN lev- els were significantly higher, the serum C-reactive protein (CRP) levels were significantly higher, and the serum BNP levels were significantly higher in the dead group than in the alive group.</p>			
Concern that the included participants and setting do not match the review question	CONCERN: (low/ high/ unclear)	Low	
<p><i>Rationale of applicability rating: The characteristics of the study population are consistent with the criteria of the research question.</i></p>			

DOMAIN 2: Predictors			
A. Risk of Bias			
<p><i>List and describe predictors included in the final model, e.g. definition and timing of assessment:</i> We constructed a predictive scoring sys- tem based on the significant APACHE II factors in the alive group (mean BP, pulse, sodium, potassium, creatinine, hemat- ocrit, age, and GCS; giving one point based on each cut-off value), defined as the APACHE-HF score (The cut-off values for each factor were defined by ROC curves as follows: mean BP (91.5mmHg, AUC= 0.678), pulse (110.5 beats/min, AUC= 0.594), sodium (137.5 mmol/L, AUC= 0.613), potassium (4.85 mmol/L, AUC= 0.601), hematocrit (36.95 mg/dL, AUC= 0.617), creatinine (1.475 mg/dL, AUC= 0.676), age (71.5 years, AUC= 0.572), and GCS (13.5, AUC= 0.567). We constructed a predictive scoring system based on the significant APACHE II factors in the alive group (mean BP, sodium, potassium, creatinine, age, and GCS; giving points based on the APACHE II system), defined as the Modified APACHE II score (Fig. 1B).</p>			
		De v	Val
2.1 Were predictors defined and assessed in a similar way for all participants?		Y	
2.2 Were predictor assessments made without knowledge of outcome data?		Y	
2.3 Are all predictors available at the time the model is intended to be used?		Y	
Risk of bias introduced by predictors or their assessment	RISK: (low/ high/ unclear)	Low	
<p><i>Rationale of bias rating: It has a complete description of the list of selected predictor variables, as well as the way in which they were studied in patients.</i></p>			
B. Applicability			

Concern that the definition, assessment or timing of predictors in the model do not match the review question	CONCERN: (low/ high/ unclear)	Low	Low
<i>Rationale of applicability rating:</i> <i>The evaluation time of the predictive variables and the applicability of these are consistent with the research question of the recent systematic review.</i>			
DOMAIN 3: Outcome			
A. Risk of Bias			
<i>Describe the outcome, how it was defined and determined, and the time interval between predictor assessment and outcome determination:</i> Predict adverse outcomes in patients with AHF at a 90-day follow-up. the mid-term prognosis was evaluated in terms of all-cause death and HF events defined as all-cause death or read- mission due to HF. The patients were clinically followed up in a routine outpatient clinic. For the patients followed up at other institutes, the final prognosis was determined via telephone contact.			
		Dev	Val
3.1 Was the outcome determined appropriately?		Y	
3.2 Was a pre-specified or standard outcome definition used?		Y	
3.3 Were predictors excluded from the outcome definition?		Y	
3.4 Was the outcome defined and determined in a similar way for all participants?		Y	
3.5 Was the outcome determined without knowledge of predictor information?		Y	
3.6 Was the time interval between predictor assessment and outcome determination appropriate?		Y	
Risk of bias introduced by the outcome or its determination	RISK: (low/ high/ unclear)	Low	
<i>Rationale of bias rating: Adequate description of the outcomes and the manner in which they were collected.</i>			
B. Applicability			
<i>At what time point was the outcome determined:</i> <i>It was determined during the follow-up in the future.</i>			
<i>If a composite outcome was used, describe the relative frequency/distribution of each contributing outcome:</i>			
Concern that the outcome, its definition, timing or determination do not match the review question	CONCERN: (low/ high/ unclear)	Low	
<i>Rationale of applicability rating: The time in which the outcome is determined is consistent with the proposed research question.</i>			
DOMAIN 4: Analysis			
Risk of Bias			
<i>Describe numbers of participants, number of candidate predictors (for DEV only), outcome events and events per candidate predictor (for DEV only):</i> AHF patients were divided into two groups according to in- hospital mortality :the alive group (n = 750) and the dead group (n = 74). The total number of events in the study is not mentioned.			
<i>Describe how the model was developed (predictor selection, optimism, risk groups, model performance):</i> AHF patients were divided into two groups according to in- hospital mortality: the alive group(n = 750) and the dead group (n = 74). We established two new scoring systems for AHF comparing the two groups. First , we compared the APACHE II score between the two groups using a univariate analysis. Factors associated with significantly more points in the alive group were selected to construct the new scoring system. The total score of the significant factors in the alive group was defined as the Modified APACHEII score .Second ,we compared the data for APACHE II factors between the two groups using a univariate analysis. We then scored the total			

<p>number of factors specific to survival discharge (giving one point for each factor), defined as the APACHE-HF score. The cut-off value for each factor to give one point was defined by the receiver- operating characteristic (ROC) curves for the in-hospital mortality of each factor. We determined the scores (APACHE II, Modified APACHE II, and APACHE-HF) in each patient based on data obtained at admission, according to a previous report, and evaluated the sensitivity, specificity, and positive and negative predictive value for differentiating the alive and dead groups. ROC curves were calculated to predict the optimal cut-off values. Furthermore, the mid-term prognosis was evaluated in terms of all-cause death and HF events defined as all-cause death or readmission due to HF. The patients were clinically followed up in a routine out patient clinic. For the patients followed up at other institutes, the final prognosis was determined via telephone contact. The patients were assigned to another two groups according to the cut-off values of the ROC curves for the APACHE II, Modified APACHE II, and APACHE-HF scores. The survival rates were then analyzed using Kaplan–Meier curves according to the APACHE II, Modified APACHE II, and APACHE-HF scores. The prognostic value of the APACHE-HF score in these groups compared with that observed in the group with lowest point as the referent was assessed using a Cox regression hazard model.</p>		
<p><i>Describe whether and how the model was validated, either internally (e.g. bootstrapping, cross validation, random split sample) or externally (e.g. temporal validation, geographical validation, different setting, different type of participants): The article mentions the use of a derivation cohort and a validation cohort for internal and external validation respectively.</i></p>		
<p><i>Describe the performance measures of the model, e.g. (re)calibration, discrimination, (re)classification, net benefit:</i> The APACHEII demonstrated an optimal balance between sensitivity and specificity (47.3% and 67.3% ; AUC=0.588) at 17.5 points, selecting 9.5 points for the Modified APACHEII score as a cut-off value (sensitivity 51.4%, specificity 69.4% , AUC 0.590). Meanwhile, a cut-off value of 2.5 points for the APACHE-HF score produced an optimal balance (sensitivity 87.8%, specificity 63.9%, AUC 0.779). A higher optimal balance was observed in the APACHE-HF score than in the APACHE II or Modified APACHE II scores. Based on these cut-off values, the patients were divided into low or high groups for each scoring system. The Kaplan–Meier survival curves showed that the prognosis, including all-cause death, did not differ between the patients with an APACHE II score of ≤ 17 and those with an APACHE II score of ≥ 18. Meanwhile, the prognosis was significantly poorer in the patients with a Modified APACHE II score of ≥ 10 than in those with a Modified APACHE II score of ≤ 9 and in the patients with an APACHE-HF score of ≥ 3 than in those with an APACHEII-HF score of ≤ 2. The Kaplan–Meier survival curves showed that the prognosis, including HF events, was significantly poorer in the patients with an APACHE II score of ≥ 18, a Modified APACHE II score of ≥ 10, and an APACHE-HF score of ≥ 3 than in those with an APACHE II score of ≤ 17, a Modified APACHE II of ≤ 9, and an APACHE-HF score of ≤ 2, respectively. The patients were assigned to five groups based on the APACHE- HF score [Group1: point0 (n = 70), Group2: points 1 and 2 (n = 343), Group 3: points 3 and 4 (n = 294), Group 4: points 5 and 6 (n = 106) and Group5: points 7 and 8 (n = 11)]. The multivariate Cox regression model identified belonging to Group3 (HR: 7.700, 95% CI 0.935 – 63.421), Group4 (HR: 12.357, 95% CI 1.388 – 110.010), or Group5 (HR: 18.361, 95% CI 1.478–228.145) as independent predictors of 90 – day mortality. Further more ,the multivariate Cox regression model identified belonging to Group2 (HR: 3.253, 95% CI 1.173–9.019), Group 3 (HR: 5.298, 95% CI 1.842–15.238), Group4 (HR: 6.201, 95% CI 2.040–18.846), or Group5 (HR: 9.413, 95% CI 2.432–36.436) as independent predictors of HF events during the 90 -day follow-up period. The Kaplan–Meier curves revealed a poorer prognosis, including all-cause death and HF events, in Group5 and Group4 than in the other groups, in Group 3 than in Group 2 or Group 1, and in Group 2 than in Group1.</p>		
<p><i>Describe any participants who were excluded from the analysis:</i> Patients with HF caused by acute coronary syn- drome were excluded from the study.</p>		
<p><i>Describe missing data on predictors and outcomes as well as methods used for missing data: The authors don't mention how the missing data were handled.</i></p>		
	Dev	Val
4.1 Were there a reasonable number of participants with the outcome?	N	
4.2 Were continuous and categorical predictors handled appropriately?	PY	
4.3 Were all enrolled participants included in the analysis?	N	
4.4 Were participants with missing data handled appropriately?	NI	

4.5 Was selection of predictors based on univariable analysis avoided?	Y	
4.6 Were complexities in the data (e.g. censoring, competing risks, sampling of controls) accounted for appropriately?	Y	
4.7 Were relevant model performance measures evaluated appropriately?	PY	
4.8 Were model overfitting and optimism in model performance accounted for?	NI	
4.9 Do predictors and their assigned weights in the final model correspond to the results from the reported multivariable analysis?	Y	
Risk of bias introduced by the analysis	RISK: (low/ high/ unclear)	Low
<i>Rationale of bias rating: An adequate analysis of the predictive model was made, the outcome and performance coincides with those sought in the systematic review.</i>		

Step 4: Overall assessment

Use the following tables to reach overall judgments about risk of bias and concerns for applicability of the prediction model evaluation (development and/or validation) across all assessed domains.

Reaching an overall judgment about risk of bias of the prediction model evaluation	
Low risk of bias	If all domains were rated low risk of bias. If a <u>prediction model was developed without any external validation</u> , and it was rated as <u>low risk of bias for all domains</u> , consider downgrading to high risk of bias . Such a model can only be considered as low risk of bias, if the development was based on a very large data set and included some form of internal validation.
High risk of bias	If at least one domain is judged to be at high risk of bias .
Unclear risk of bias	If an unclear risk of bias was noted in at least one domain and it was low risk for all other domains.

Complete for each evaluation of a distinct model.

Reaching an overall judgment about applicability of the prediction model evaluation	
Low concerns for applicability	If low concerns for applicability for all domains, the prediction model evaluation is judged to have low concerns for applicability .
High concerns for applicability	If high concerns for applicability for at least one domain, the prediction model evaluation is judged to have high concerns for applicability .
Unclear concerns for applicability	If unclear concerns (but no “high concern”) for applicability for at least one domain, the prediction model evaluation is judged to have unclear concerns for applicability overall.

Overall judgment about risk of bias and applicability of the prediction model evaluation		
Overall judgment of risk of bias	RISK: (low/ high/ unclear)	
<i>Summary of sources of potential bias: Low</i>		

Overall judgment of applicability	CONCERN: (low/ high/ unclear)	
Summary of applicability concerns: Low		

PROBAST = Prediction model Risk Of Bias ASsessment Tool.

29. Parenica J, et al:

Step 2: Classify the type of prediction model evaluation

Classify the evaluation based on its aim			
Type of prediction study	PROBAST boxes to complete	Tick as appropriate	Definitions for type of prediction model study
Development only	Dev		Prediction model development without external validation. These studies may include internal validation methods such as bootstrapping and cross-validation techniques
Development and validation	Dev and Val		Prediction model development combined with external validation in other participants in the same article
Validation only	Val	X	External validation of existing (previously developed) model in other participants
<i>This table should be completed once for each publication being assessed and for each relevant outcome in your review.</i>			
Publication reference	Parenica J, Kala P, Goldbergova M, Tomandl J, Spinar J, Littnerova S, et al. Natriuretic peptides, nitrite/nitrate and superoxide dismutase have additional value on top of the GRACE score in prediction of one-year mortality and rehospitalisation for heart failure in STEMI patients — Multiple biomarkers prospective cohort study. International Journal of Cardiology.2016;211:96–104.		
Models of interest	GRACE–natriuretic peptide (NP) score		
Outcome of interest	The aim of our prospective observational cohort study was to evaluate the prognostic value of biomarkers of different pathophysiological pathways for the occurrence of 1-year all-cause mortality and hospitalisation due to acute heart failure.		

Step 3: Assess risk of bias and applicability

DOMAIN 1: Participants
A. Risk of Bias
<i>Describe the sources of data and criteria for participant selection: From November 2005 to October 2008, 913 patients with STEMI were referred for primary PCI. They were admitted to the Coronary Care Unit, Department of Cardiology, University Hospital Brno. Diagnosis of STEMI was based on MI symptoms, appropriate ECG changes [ST-segment elevation at the J point in two contiguous leads with cut-off points ≥ 0.2 mV in leads V1–V3 and ≥ 0.1 mV in the other leads, or new left bundle branch block and elevation of levels of cardiac troponin</i>

<p>I (marker of myocardial necrosis)]. The time from the onset of chest pain to primary PCI was 12 h. The diagnosis of AHF was based on clinical signs upon hospital admission, and then during hospitalisation (Killip II–IV). The exclusion criteria were: age ≥ 80 years ($n = 112$); known or newly diagnosed malignancy; inflammatory disease, connective-tissue disease, or disease other than cardiovascular that would clearly limit the one-year prognosis ($n = 99$); refusal to sign the informed consent, or a supposed non-compliance ($n = 74$); geographic reasons (distance from the place of residence to the hospital ≥ 100 km, without the possibility of follow-up; $n = 35$). A total of 593 patients with STEMI were included. All patients were Caucasian. The standard recommended medical therapy included dual antiplatelet therapy, angiotensin-converting enzyme inhibitors, betablockers, and statins.</p>			
		De v	Val
1.1 Were appropriate data sources used, e.g. cohort, RCT or nested case-control study data?		Y	
1.2 Were all inclusions and exclusions of participants appropriate?		Y	
Risk of bias introduced by selection of participants	RISK: (low/ high/ unclear)	Low	
Rationale of bias rating: Applied correct criteria for inclusion of participants			
B. Applicability			
Describe included participants, setting and dates: A cohort for validation of our combined GRACE–natriuretic peptide model included 667 patients with STEMI treated by primary PCI. The patients were enrolled from June 2009 to January 2013. All patients were followed up for one year; mortality, hospitalisation for AHF, reinfarctions and TIA/strokes were recorded. BNP was evaluated in all patients 24 h after the onset of STEMI.			
Concern that the included participants and setting do not match the review question	CONCERN: (low/ high/ unclear)	Low	
Rationale of applicability rating: The characteristics of the study population are consistent with the criteria of the research question.			

DOMAIN 2: Predictors			
A. Risk of Bias			
List and describe predictors included in the final model, e.g. definition and timing of assessment: clinical score GRACE (based on age, heart rate, systolic blood pressure, creatinine, cardiac arrest at admission, ST segment deviation on ECG, elevated cardiac enzymes and signs of heart failure)			
		De v	Val
2.1 Were predictors defined and assessed in a similar way for all participants?		Y	
2.2 Were predictor assessments made without knowledge of outcome data?		Y	
2.3 Are all predictors available at the time the model is intended to be used?		Y	
Risk of bias introduced by predictors or their assessment	RISK: (low/ high/ unclear)	Low	
Rationale of bias rating: They adequately describe the way of choosing final predictors.			
B. Applicability			
Concern that the definition, assessment or timing of predictors in the model do not match the review question	CONCERN: (low/ high/ unclear)	Low	
Rationale of applicability rating: The evaluation time of the predictive variables and the applicability of these are consistent with the research question of the recent systematic review.			

DOMAIN 3: Outcome			
A. Risk of Bias			
<p><i>Describe the outcome, how it was defined and determined, and the time interval between predictor assessment and outcome determination:</i> The primary endpoint was the combination of one-year all-cause mortality and a new or worsening pre-existing HF requiring hospitalisation in an acute-care facility, primarily for the treatment of HF; secondary endpoints involved all-cause mortality, hospitalisation for AHF, re-infarction (according to the Third Universal Definition of Myocardial Infarction), and transient ischaemic attack (TIA; a transient episode of focal neurological dysfunction caused by brain, spinal cord, or retinal ischaemia, without acute infarction) or stroke. All endpoints were adjudicated by two experienced cardiologists. The definitions of used end-points are consistent with the published document 2014 ACC/AHA Key Data Elements and Definitions for Cardiovascular Endpoint Events in Clinical Trials.</p>			
		Dev	Val
3.1 Was the outcome determined appropriately?		Y	
3.2 Was a pre-specified or standard outcome definition used?		Y	
3.3 Were predictors excluded from the outcome definition?		Y	
3.4 Was the outcome defined and determined in a similar way for all participants?		Y	
3.5 Was the outcome determined without knowledge of predictor information?		Y	
3.6 Was the time interval between predictor assessment and outcome determination appropriate?		Y	
Risk of bias introduced by the outcome or its determination	RISK: (low/ high/ unclear)	Low	
<p><i>Rationale of bias rating:</i> They show an outcome that matches the question of the systematic review..</p>			
B. Applicability			
<p><i>At what time point was the outcome determined:</i> It was determined during the follow-up in the future.</p> <p><i>If a composite outcome was used, describe the relative frequency/distribution of each contributing outcome:</i></p>			
Concern that the outcome, its definition, timing or determination do not match the review question	CONCERN: (low/ high/ unclear)	Low	
<p><i>Rationale of applicability rating:</i> The time in which the outcome is determined is consistent with the proposed research question.</p>			
DOMAIN 4: Analysis			
Risk of Bias			
<p><i>Describe numbers of participants, number of candidate predictors (for DEV only), outcome events and events per candidate predictor (for DEV only):</i> The one-year mortality was 6.7% (40 patients); 15 patients (2.5%) were hospitalised for AHF, and 55 of 593 patients (9.2%) had primary endpoints at 1 year. the number of candidate variables isn't mentioned</p>			
<p><i>Describe how the model was developed (predictor selection, optimism, risk groups, model performance):</i> Standard descriptive statistics were applied in the analysis: absolute and relative frequencies for categorical variables and median supplemented by 5th–95th percentile ranges for quantitative variables. The value of GRACE risk score, which is derived from several variables (age, heart rate, systolic blood pressure, initial serum creatinine concentration, congestive heart failure — Killip class, cardiac arrest on admission, ST segment deviation and elevated cardiac markers), was calculated separately for each patient according to the published algorithm. The significance of differences among groups of patients was analysed using the maximum-likelihood χ^2 test for categorical variables and Mann–Whitney or Kruskal–Wallis tests for quantitative variables. The statistical significance was adjusted by Bonferroni correction. The relationship between quantitative variables was assessed using the Spearman correlation coefficient. Receiver operating curve (ROC) analysis was adopted to describe the predictive power of variables and identification of optimal cut-off</p>			

<p>points. Reclassification analyses [net reclassification improvement (NRI), continuous net reclassification improvement (cNRI) and integrated discrimination improvement (IDI)] were used for the characterisation of improvement of the standard GRACE risk score by the addition of new markers. Analyses were computed using SPSS v22 (IBM Corporation, Armonk, NY, USA) and R 2.15 with library PredictABEL. $\alpha = 0.05$ was considered significant in all analyses.</p>			
<p><i>Describe whether and how the model was validated, either internally (e.g. bootstrapping, cross validation, random split sample) or externally (e.g. temporal validation, geographical validation, different setting, different type of participants): The model was used in a different population from the development study. In addition, this time was used for outcomes corresponding to AHF and not acute coronary syndrome.</i></p>			
<p><i>Describe the performance measures of the model, e.g. (re)calibration, discrimination, (re)classification, net benefit: We tested the ability of the selected biomarkers and GRACE to predict the secondary endpoints within one year. C statistics demonstrated a moderately strong ability of GRACE, BNP, NT-proBNP, SOD and NOx to predict the one-year mortality (AUC was 0.852, resp. 0.845, resp. 0.852, resp. 0.792, resp. 0.727, for all $p < 0.001$) and a moderately strong ability to predict hospitalisation due to AHF (AUC was 0.732, resp. 0.819, resp. 0.795, resp. 0.779, resp. 0.761, for all $p < 0.01$), but there was no ability to predict the onset of re-infarction or TIA/stroke. We did not find any significant relationship between any from all evaluated biomarkers and re-infarction or TIA/stroke (data not shown). According to C statistics, GRACE score alone, BNP and NT-proBNP demonstrated a good ability to predict the occurrence of death and/or hospitalisation due to AHF during 2 and 3 years (AUC for 2 years 0.791, 0.810, and 0.795 respectively, and AUC for 3 years 0.747, 0.757, and 0.735 respectively, for all $p < 0.001$). SOD and NOx also demonstrated a significant but weaker ability to predict the primary endpoints (AUC for 2 years 0.758 and 0.712 respectively, AUC for 3 years 0.686 and 0.630 respectively, for $p < 0.001$). In another validation prospective cohort of 667 STEMI patients treated by primary PCI, we have tested an additional prognostic value of BNP to GRACE score. According to the ROC analysis, AUC for the prediction of one-year mortality and/or hospitalisation for AHF was 0.756 (95% CI 0.679; 0.833; $p < 0.001$) for the GRACE score, and 0.749 (0.667; 0.831; $p < 0.001$) for BNP. An addition of BNP to the GRACE score demonstrated a significant increase in AUC to 0.793 (0.720; 0.865, $p < 0.001$) in comparison to GRACE score alone, according to the DeLong's test. The combination of GRACE score ≥ 150 and BNP ≥ 330 pg/ml characterises a group of high-risk patients. In comparison with the above-cited papers, our results demonstrate high values of cNRI, particularly for BNP, NTproBNP and SOD (72.1–106.6%).</i></p>			
<p><i>Describe any participants who were excluded from the analysis:</i> The exclusion criteria were: age ≥ 80 years ($n = 112$); known or newly diagnosed malignancy; inflammatory disease, connective-tissue disease, or disease other than cardiovascular that would clearly limit the one-year prognosis ($n = 99$); refusal to sign the informed consent, or a supposed non-compliance ($n = 74$); geographic reasons (distance from the place of residence to the hospital ≥ 100 km, without the possibility of follow-up; $n = 35$).</p>			
<p><i>Describe missing data on predictors and outcomes as well as methods used for missing data: The handling of missing data is not mentioned. The handling of missing data is not mentioned.</i></p>			
		Dev	Val
4.1 Were there a reasonable number of participants with the outcome?		PN	
4.2 Were continuous and categorical predictors handled appropriately?		Y	
4.3 Were all enrolled participants included in the analysis?		Y	
4.4 Were participants with missing data handled appropriately?		NI	
4.5 Was selection of predictors based on univariable analysis avoided?		Y	
4.6 Were complexities in the data (e.g. censoring, competing risks, sampling of controls) accounted for appropriately?		NI	
4.7 Were relevant model performance measures evaluated appropriately?		Y	
4.8 Were model overfitting and optimism in model performance accounted for?		NI	
4.9 Do predictors and their assigned weights in the final model correspond to the results from the reported multivariable analysis?		Y	
Risk of bias introduced by the analysis	RISK: (low/ high/ unclear)	Low	

Rationale of bias rating: It has a good analysis, agrees with the research question of the systematic review.

Step 4: Overall assessment

Use the following tables to reach overall judgments about risk of bias and concerns for applicability of the prediction model evaluation (development and/or validation) across all assessed domains.

Reaching an overall judgment about risk of bias of the prediction model evaluation	
Low risk of bias	If all domains were rated low risk of bias. If a <u>prediction model was developed without any external validation</u> , and it was rated as <u>low risk of bias for all domains</u> , consider downgrading to high risk of bias . Such a model can only be considered as low risk of bias, if the development was based on a very large data set and included some form of internal validation.
High risk of bias	If at least one domain is judged to be at high risk of bias .
Unclear risk of bias	If an unclear risk of bias was noted in at least one domain and it was low risk for all other domains.

Complete for each evaluation of a distinct model.

Reaching an overall judgment about applicability of the prediction model evaluation	
Low concerns for applicability	If low concerns for applicability for all domains, the prediction model evaluation is judged to have low concerns for applicability .
High concerns for applicability	If high concerns for applicability for at least one domain, the prediction model evaluation is judged to have high concerns for applicability .
Unclear concerns for applicability	If unclear concerns (but no “high concern”) for applicability for at least one domain, the prediction model evaluation is judged to have unclear concerns for applicability overall.

Overall judgment about risk of bias and applicability of the prediction model evaluation		
Overall judgment of risk of bias	RISK: (low/ high/ unclear)	
<i>Summary of sources of potential bias: Low</i>		
Overall judgment of applicability	CONCERN: (low/ high/ unclear)	
<i>Summary of applicability concerns: Low</i>		

PROBAST = Prediction model Risk Of Bias ASsessment Tool.

30. Sasaki N, et al:

Step 2: Classify the type of prediction model evaluation

Classify the evaluation based on its aim			
Type of prediction study	PROBAST boxes to complete	Tick as appropriate	Definitions for type of prediction model study
Development only	Dev		Prediction model development without external validation. These studies may include internal validation methods such as bootstrapping and cross-validation techniques
Development and validation	Dev and Val	X	Prediction model development combined with external validation in other participants in the same article
Validation only	Val		External validation of existing (previously developed) model in other participants
<i>This table should be completed once for each publication being assessed and for each relevant outcome in your review.</i>			
Publication reference	Sasaki N. Lee J. Park S. Umegaki T. Kunisawa S. Development and Validation of an Acute Heart Failure-Specific Mortality Predictive Model Based on Administrative Data. Canadian Journal of Cardiology. Volume 29 (2013) 1055-1061. http://dx.doi.org/10.1016/j.cjca.2012.11.021		
Models of interest	Heart Failure-Specific Mortality Predictive Model		
Outcome of interest	To develop an accurate, practical, and reproducible risk adjustment model to predict AHF in-hospital mortality using factors identifiable from administrative data, and to apply this model to an interhospital comparison.		

Step 3: Assess risk of bias and applicability

DOMAIN 1: Participants			
A. Risk of Bias			
Describe the sources of data and criteria for participant selection: selection criteria were also used: Patients who had both an 'acute exacerbation' of heart failure code and NYHA functional class II or higher, which were available within the DPC system and patients who were at least 20 years of age at admission. Data were collected from 19,792 patients across 139 hospitals with a primary diagnosis of heart failure.			
		Dev	Val
1.1 Were appropriate data sources used, e.g. cohort, RCT or nested case-control study data?		Y	Y
1.2 Were all inclusions and exclusions of participants appropriate?		Y	Y
Risk of bias introduced by selection of participants	RISK: (low/unclear) high/	Low	Low
<i>Rationale of bias rating:</i>			
Inclusion and exclusion criteria was correctly mentioned.			
B. Applicability			

<i>Describe included participants, setting and dates: The study population consisted of patients with AHF.</i>			
Concern that the included participants and setting do not match the review question	CONCERN: (low/ high/ unclear)	Low	Low
<i>Rationale of applicability rating: Information from patients were correctly mentioned.</i>			

DOMAIN 2: Predictors			
A. Risk of Bias			
<i>List and describe predictors included in the final model, e.g. definition and timing of assessment: Female, age, admission route, NYHA, Severe respiratory failure because of AHF, Ischemic heart disease, hypertension, AF, life-threatening arrhythmia, CRF, shock.</i>			
		Dev	Val
2.1 Were predictors defined and assessed in a similar way for all participants?		Y	Y
2.2 Were predictor assessments made without knowledge of outcome data?		Y	Y
2.3 Are all predictors available at the time the model is intended to be used?		Y	Y
Risk of bias introduced by predictors or their assessment	RISK: (low/ high/ unclear)	Low	Low
<i>Rationale of bias rating: The article has a good selection of variables combining variables known as risk factors in AHF.</i>			
B. Applicability			
Concern that the definition, assessment or timing of predictors in the model do not match the review question	CONCERN: (low/ high/ unclear)	Unclear	Unclear
<i>Rationale of applicability rating: They show the models they developed, but no score</i>			
DOMAIN 3: Outcome			
A. Risk of Bias			
<i>Describe the outcome, how it was defined and determined, and the time interval between predictor assessment and outcome determination: To develop an accurate, practical, and reproducible risk adjustment model to predict AHF in-hospital mortality using factors identifiable from administrative data, and to apply this model to an interhospital comparison.</i>			
		Dev	Val
3.1 Was the outcome determined appropriately?		Y	Y
3.2 Was a pre-specified or standard outcome definition used?		Y	Y
3.3 Were predictors excluded from the outcome definition?		Y	Y
3.4 Was the outcome defined and determined in a similar way for all participants?		Y	Y
3.5 Was the outcome determined without knowledge of predictor information?		Y	Y
3.6 Was the time interval between predictor assessment and outcome determination appropriate?		Y	Y
Risk of bias introduced by the outcome or its determination	RISK: (low/ high/ unclear)	Low	Low

<i>Rationale of bias rating: It was very well defined outcome, with good model</i>			
B. Applicability			
<i>At what time point was the outcome determined: at one time in the future</i>			
<i>If a composite outcome was used, describe the relative frequency/distribution of each contributing outcome:</i>			
Concern that the outcome, its definition, timing or determination do not match the review question	CONCERN: (low/ high/ unclear)	Low	Low
<i>Rationale of applicability rating: It was well defined the outcome and it match with the review question.</i>			
DOMAIN 4: Analysis			
Risk of Bias			
<i>Describe numbers of participants, number of candidate predictors (for DEV only), outcome events and events per candidate predictor (for DEV only):</i> 8620 patients from 86 hospitals. Candidate variables are sex, age, hospital admission route, NYHA, severe respiratory failure because of AHF, Ischemic heart disease, hypertension, AF, life-threatening arrhythmia, CRF, shock. No information about EPV was presented.			
<i>Describe how the model was developed (predictor selection, optimism, risk groups, model performance):</i> Multivariable logistic regression was used to estimate the predictors of in-hospital mortality of the original full dataset. We retained all covariates in the final model. In-hospital mortality rate was used as the primary outcome measure. Two types of mortality prediction models (model 1 and model 2; using identical predictors without and with NYHA functional class, respectively) were constructed with multivariable logistic regression using the original dataset. (training set). Discrimination of the logistic regression models Clinical factors related to patient severity were defined as follows: admission route was identified using an "emergency" admission code, which allowed for the identification of unplanned admissions with or without ambulance use. Severe respiratory failure because of AHF was identified with procedure codes reflecting acute mechanical ventilation use within 48 hours postadmission. Potential comorbidities were selected from the previous published literature as risk factors was assessed using the c-statistic. To validate the prediction model, a model in the bootstrap sample dataset (n 1/4 8620) was derived and Hosmer-Lemeshow test was performed to evaluate model calibration (P > 0.05 is considered favourable)			
<i>Describe whether and how the model was validated, either internally (e.g. bootstrapping, cross validation, random split sample) or externally (e.g. temporal validation, geographical validation, different setting, different type of participants):</i> They develop the score and validate it. Bootstrapping was used for internal validation.			
<i>Describe the performance measures of the model, e.g. (re)calibration, discrimination, (re)classification, net benefit:</i> Factors that were independently associated with higher in-hospital mortality included advanced age, NYHA functional class, severe respiratory failure because of AHF, shock, life-threatening arrhythmia including ventricular fibrillation/flutter or sinus arrest, and mild to moderate chronic renal failure. In contrast, hypertension, IHD, and atrial fibrillation/flutter were associated with lower in-hospital mortality.			
The c-statistics for the training models of model 1 and model 2 were 0.77 (95% CI, 0.75-0.79) and 0.81 (95% CI, 0.79- 0.82), respectively. In the sample dataset, the discriminative ability of the model was maintained with the c-statistics at 0.76 (95% CI, 0.74-0.78) and 0.80 (95% CI, 0.78-0.82), respectively, after bootstrap			

correction. The Hosmer-Lemeshow test showed no significance (P 1/4 0.44 and P 1/4 0.88, respectively), indicating good calibration.			
Describe any participants who were excluded from the analysis: Patients were excluded from the analysis if they simultaneously had acute myocardial infarction or if they had other conditions indistinguishable from AHF (n 1/4 2211), including cardiopulmonary arrest (ICD-10 codes: I46.1, I46.9, R96), acute respiratory distress syndrome (ICD-10 code: J80), severe pneumonia (ICD-10 codes: J10.0, J11.0, J12-J18, J69), pleuritis (ICD-10 codes: A15.6, A16.5, R09.1, J90, J91, J94.x), and severe renal failure (ICD-10 code: N18.0) with or without dialysis at admission. Patients with a length of stay longer than 3 standard deviations from the mean (n 1/4 169) and an invalid mortality record (n 1/4 1) were also excluded from the analysis.			
Describe missing data on predictors and outcomes as well as methods used for missing data: Do not present any missing data			
		Dev	Val
4.1 Were there a reasonable number of participants with the outcome?		Y	Y
4.2 Were continuous and categorical predictors handled appropriately?		Y	Y
4.3 Were all enrolled participants included in the analysis?		Y	Y
4.4 Were participants with missing data handled appropriately?		Y	Y
4.5 Was selection of predictors based on univariable analysis avoided?		Y	Y
4.6 Were complexities in the data (e.g. censoring, competing risks, sampling of controls) accounted for appropriately?		NI	NI
4.7 Were relevant model performance measures evaluated appropriately?		Y	Y
4.8 Were model overfitting and optimism in model performance accounted for?		Y	Y
4.9 Do predictors and their assigned weights in the final model correspond to the results from the reported multivariable analysis?		Y	Y
Risk of bias introduced by the analysis	RISK: (low/ high/ unclear)	Low	Low
Rationale of bias rating: Describe how they get the score			

Step 4: Overall assessment

Use the following tables to reach overall judgments about risk of bias and concerns for applicability of the prediction model evaluation (development and/or validation) across all assessed domains.

Reaching an overall judgment about risk of bias of the prediction model evaluation	
Low risk of bias	If all domains were rated low risk of bias. If a prediction model was developed without any external validation, and it was rated as <u>low risk of bias for all domains</u> , consider downgrading to high risk of bias . Such a model can only be considered as low risk of bias, if the development was based on a very large data set and included some form of internal validation.
High risk of bias	If at least one domain is judged to be at high risk of bias .
Unclear risk of bias	If an unclear risk of bias was noted in at least one domain and it was low risk for all other domains.

Complete for each evaluation of a distinct model.

Reaching an overall judgment about applicability of the prediction model evaluation

Low concerns for applicability	If low concerns for applicability for all domains, the prediction model evaluation is judged to have low concerns for applicability .
High concerns for applicability	If high concerns for applicability for at least one domain, the prediction model evaluation is judged to have high concerns for applicability .
Unclear concerns for applicability	If unclear concerns (but no “high concern”) for applicability for at least one domain, the prediction model evaluation is judged to have unclear concerns for applicability overall.

Overall judgment about risk of bias and applicability of the prediction model evaluation		
Overall judgment of risk of bias	RISK: (low/ high/ unclear)	Low
<i>Summary of sources of potential bias: Low</i>		
Overall judgment of applicability	CONCERN: (low/ high/ unclear)	High
<i>Summary of applicability concerns: High</i>		

PROBAST = Prediction model Risk Of Bias ASsessment

31. Scrutinio D, et al. (2013)

Step 2: Classify the type of prediction model evaluation

Classify the evaluation based on its aim			
Type of prediction study	PROBAST boxes to complete	Tick as appropriate	Definitions for type of prediction model study
Development only	Dev		Prediction model development without external validation. These studies may include internal validation methods such as bootstrapping and cross-validation techniques
Development and	Dev and Val	X	Prediction model development combined with external validation in other participants in the same

validation			article
Validation only	Val		External validation of existing (previously developed) model in other participants
<i>This table should be completed once for each publication being assessed and for each relevant outcome in your review.</i>			
Publication reference	Scrutinio D, Ammirati E, Guida P, Passantino A, Raimondo R. Clinical utility of N-terminal pro-B-type natriuretic peptide for risk stratification of patients with acute decompensated heart failure. Derivation and validation of the ADHF/NT-proBNP risk score. International Journal of Cardiology 168 (2013) 2120–2126. http://dx.doi.org/10.1016/j.ijcard.2013.01.005		
Models of interest	ADHF/NT-proBNP risk score		
Outcome of interest	Determine whether NT- proBNP can improve reclassification of ADHF patients into appropriate risk strata and to develop and validate a point-based risk score for 1-year mortality.		

Step 3: Assess risk of bias and applicability

DOMAIN 1: Participants			
A. Risk of Bias			
Describe the sources of data and criteria for participant selection: 541 patients admitted for decompensation of chronic, established HF with symptoms sufficient to warrant hospitalization to the Cardiology Division of S. Maugeri Foundation (Bari, Italy) from April 2005 to February 2011 were retrospectively identified using a computer-generated list obtained from our administrative database and reviewing electronic and paper medical records.			
		De v	Val
1.1 Were appropriate data sources used, e.g. cohort, RCT or nested case-control study data?		Y	Y
1.2 Were all inclusions and exclusions of participants appropriate?		Y	Y
Risk of bias introduced by selection of participants	RISK: (low/ high/ unclear)	Low	Low
<i>Rationale of bias rating:</i>			
Inclusion criteria was correctly mentioned.			
B. Applicability			
Describe included participants, setting and dates: The study population consisted of patients with AHF.			
Concern that the included participants and setting do not match the review question	CONCERN: (low/ high/ unclear)	Low	Low
<i>Rationale of applicability rating: Information from patients were correctly mentioned.</i>			

DOMAIN 2: Predictors		
A. Risk of Bias		
List and describe predictors included in the final model, e.g. definition and timing of assessment: COPD, SBP <100, eGFR, Sodium, Hb, NT-proBNP, LVEF, TR.		
	De v	Val
2.1 Were predictors defined and assessed in a similar way for all participants?	Y	Y
2.2 Were predictor assessments made without knowledge of outcome data?	Y	Y

2.3 Are all predictors available at the time the model is intended to be used?		Y	Y
Risk of bias introduced by predictors or their assessment	RISK: (low/ high/ unclear)	Low	Low
<i>Rationale of bias rating: The article has a good selection of variables combining variables known as risk factors in ADHF.</i>			
B. Applicability			
Concern that the definition, assessment or timing of predictors in the model do not match the review question	CONCERN: (low/ high/ unclear)	Low	Low
<i>Rationale of applicability rating: They use a simple group of variables, easy to get for develop one usefull and very applicable model.</i>			
DOMAIN 3: Outcome			
A. Risk of Bias			
<i>Describe the outcome, how it was defined and determined, and the time interval between predictor assessment and outcome determination: Determine whether NT- proBNP can improve reclassification of ADHF patients into appropriate risk strata and to develop and validate a point-based risk score for 1-year mortality.</i>			
		Dev	Val
3.1 Was the outcome determined appropriately?		Y	Y
3.2 Was a pre-specified or standard outcome definition used?		Y	Y
3.3 Were predictors excluded from the outcome definition?		Y	N
3.4 Was the outcome defined and determined in a similar way for all participants?		Y	Y
3.5 Was the outcome determined without knowledge of predictor information?		Y	Y
3.6 Was the time interval between predictor assessment and outcome determination appropriate?		Y	Y
Risk of bias introduced by the outcome or its determination	RISK: (low/ high/ unclear)	Low	Low
<i>Rationale of bias rating: It was very well defined outcome, with good model</i>			
B. Applicability			
<i>At what time point was the outcome determined: at one time in the future</i>			
<i>If a composite outcome was used, describe the relative frequency/distribution of each contributing outcome:</i>			
Concern that the outcome, its definition, timing or determination do not match the review question	CONCERN: (low/ high/ unclear)	Low	Low
<i>Rationale of applicability rating: The outcome and its definition fit with the review question</i>			
DOMAIN 4: Analysis			
Risk of Bias			
<i>Describe numbers of participants, number of candidate predictors (for DEV only), outcome events and events per candidate predictor (for DEV only): There were 454 patients in the derivation cohort and 392 patients in the validation cohort. No candidate predictors or EPV.</i>			

Describe how the model was developed (predictor selection, optimism, risk groups, model performance): Score-based prediction rule was developed from logistic regression model by using an integer-based scoring system. The variables selected in the multivariable NT-proBNP algorithm were used to build the risk score.			
Describe whether and how the model was validated, either internally (e.g. bootstrapping, cross validation, random split sample) or externally (e.g. temporal validation, geographical validation, different setting, different type of participants): They develop the score and validated externally.			
Describe the performance measures of the model, e.g. (re)calibration, discrimination, (re)classification, net benefit: the C index of the logistic regression model used to develop the risk score was 0.839 (95% CI: 0.798–0.880) and the Hosmer–Lemeshow statistic 1.23 ($p = 0.542$), indicating that the model was well calibrated. Without NT-proBNP, the C index was 0.825 (95% CI: 0.783–0.867). In the validation cohort, the C index was 0.768 (95% CI: 0.711–0.817) and the Hosmer–Lemeshow. ($p=0.001$), indicating that the model did not calibrate perfectly. Without NT-proBNP, the C index was 0.760 (95% CI: 0.704–0.816).			
The C index of the risk score model with continuous values for covariates rather than categorized versions was 0.854 in the derivation and 0.794 in the validation cohort. When integer categories rather than continuous values for covariates were used in the risk reclassification analysis, the NRI was no longer statistically significant while the IDI was 0.035 ($p=0.0006$). Because of the diminished calibration in the validation cohort, we recalibrated the regression model. This resulted in improved calibration (Hosmer–Lemeshow statistic: 2.76; $p = 0.251$). The C index of the ADHF/NT-proBNP risk score to predict the combined outcome of mortality, UHT, and LVAD implantation in the external validation cohort was 0.72 (95% CI: 0.66–0.77) and the Hosmer–Lemeshow statistic.			
Describe any participants who were excluded from the analysis: had no NT-proBNP recorded at admission ($n=33$) and those with isolated right ventricular HF ($n=18$), acute coronary syndromes or angina pectoris ($n=14$), hypertrophic cardiomyopathy ($n=10$), or recent cardiac surgical or percutaneous procedures ($n=12$).			
Describe missing data on predictors and outcomes as well as methods used for missing data: Data were >99.6% complete in both the derivation (0.1% missing) and validation (0.3% missing) cohort.			
		Dev	Val
4.1 Were there a reasonable number of participants with the outcome?		Y	Y
4.2 Were continuous and categorical predictors handled appropriately?		Y	Y
4.3 Were all enrolled participants included in the analysis?		Y	Y
4.4 Were participants with missing data handled appropriately?		NI	NI
4.5 Was selection of predictors based on univariable analysis avoided?		Y	NI
4.6 Were complexities in the data (e.g. censoring, competing risks, sampling of controls) accounted for appropriately?		NI	NI
4.7 Were relevant model performance measures evaluated appropriately?		Y	Y
4.8 Were model overfitting and optimism in model performance accounted for?		Y	Y
4.9 Do predictors and their assigned weights in the final model correspond to the results from the reported multivariable analysis?		Y	Y
Risk of bias introduced by the analysis	RISK: (low/ high/ unclear)	Low	Low
Rationale of bias rating: Describe how they get the score			

Step 4: Overall assessment

Use the following tables to reach overall judgments about risk of bias and concerns for applicability of the prediction model evaluation (development and/or validation) across all assessed domains.

Reaching an overall judgment about risk of bias of the prediction model evaluation	
Low risk of bias	If all domains were rated low risk of bias. If a prediction model was developed without any external validation, and it was rated as <u>low risk of bias for all domains</u> , consider downgrading to high risk of bias . Such a model can only be considered as low risk of bias, if the development was based on a very large data set and included some form of internal validation.
High risk of bias	If at least one domain is judged to be at high risk of bias .
Unclear risk of bias	If an unclear risk of bias was noted in at least one domain and it was low risk for all other domains.

Complete for each evaluation of a distinct model.

Reaching an overall judgment about applicability of the prediction model evaluation	
Low concerns for applicability	If low concerns for applicability for all domains, the prediction model evaluation is judged to have low concerns for applicability .
High concerns for applicability	If high concerns for applicability for at least one domain, the prediction model evaluation is judged to have high concerns for applicability .
Unclear concerns for applicability	If unclear concerns (but no “high concern”) for applicability for at least one domain, the prediction model evaluation is judged to have unclear concerns for applicability overall.

Overall judgment about risk of bias and applicability of the prediction model evaluation		
Overall judgment of risk of bias	RISK: (low/ high/ unclear)	Low
<i>Summary of sources of potential bias: Low</i>		
Overall judgment of applicability	CONCERN: (low/ high/ unclear)	Low
<i>Summary of applicability concerns: Low</i>		

PROBAST = Prediction model Risk Of Bias ASsessment Tool.

32. Scrutinio D, et al. (2014)

Step 2: Classify the type of prediction model evaluation

Classify the evaluation based on its aim			
Type of prediction study	PROBAST boxes to complete	Tick as appropriate	Definitions for type of prediction model study
Development only	Dev		Prediction model development without external validation. These studies may include internal validation methods such as bootstrapping and cross-validation techniques
Development and validation	Dev and Val		Prediction model development combined with external validation in other participants in the same article
Validation only	Val	X	External validation of existing (previously developed) model in other participants
<i>This table should be completed once for each publication being assessed and for each relevant outcome in your review.</i>			
Publication reference	Scrutinio D, Ammirati E, Guida P, Passantino A, Raimondo R. The ADHF/NT-proBNP risk score to predict 1-year mortality in hospitalized patients with advanced decompensated heart failure. The Journal of Heart and Lung Transplantation. Vol 33, No 4, April 2014. http://dx.doi.org/10.1016/j.healun.2013.12.005		
Models of interest	ADHF/NT-proBNP risk score		
Outcome of interest	Assess the performance of the ADHF/NT-proBNP risk score to predict death in hospitalized patients with AHF, accounting for the competing events VAD implantation and HT.		

Step 3: Assess risk of bias and applicability

DOMAIN 1: Participants			
A. Risk of Bias			
Describe the sources of data and criteria for participant selection: The study population comprised 445 patients at admission. selection criteria: current hospitalization for worsening of chronic established HF, history of heart failure for at least 1 year, receiving chronic treatment with standard therapies, New York Heart Association Functional Class III/IV symptoms and evidence of severe left ventricular systolic dysfunction (left ventricular ejection fraction [LVEF] ≤ 0.30 as measured by two-dimensional echocardiography) at admission, and need for intravenous diuretic and/or inotropic treatment.			
		Dev	Val
1.1 Were appropriate data sources used, e.g. cohort, RCT or nested case-control study data?			PY
1.2 Were all inclusions and exclusions of participants appropriate?			Y
Risk of bias introduced by selection of participants	RISK: (low/ high/ unclear)		Low
<i>Rationale of bias rating:</i>			
Inclusion criteria and exclusion criteria were well presented. Nevertheless, no known where population came from.			
B. Applicability			

<p><i>Describe included participants, setting and dates:</i> Of the 454 patients constituting the study population, 9 (2%) were lost to follow-up, leaving 445 patients available for analysis. Mean age was 62 years, 59.6% had had Z1 hospitalizations for HF within the 6 months preceding the index admission, 78% had an implanted ICD, mean New York Heart Association class was 3.4, mean LVEF was 22.9%, and 41.3% required intravenous inotropic therapy. Of the 445 patients, 144 (32.4%) died, and 69 (15.5%) Under went VAD implantation (n = 20) or HT(n = 49) as a first event within 12 months. Death represented 67.6% of the events that occurred ≤ 1 year. Of the 322 patients aged ≤ 70 years, 76 (23.6%) died and 69 (21.4%) under went VAD implantation (n = 20) or HT (n = 49) as a first event within 12 months. Death represented 52.4% of the events that occurred ≤ 1 year.</p>			
Concern that the included participants and setting do not match the review question	CONCERN: (low/ high/ unclear)		Low
<i>Rationale of applicability rating: Information from patients were correctly mentioned.</i>			

DOMAIN 2: Predictors			
A. Risk of Bias			
<i>List and describe predictors included in the final model, e.g. definition and timing of assessment: They do not mentioned any predictors included.</i>			
		De v	Val
2.1 Were predictors defined and assessed in a similar way for all participants?			NI
2.2 Were predictor assessments made without knowledge of outcome data?			NI
2.3 Are all predictors available at the time the model is intended to be used?			NI
Risk of bias introduced by predictors or their assessment	RISK: (low/ high/ unclear)		High
<i>Rationale of bias rating: They do not mentioned any predictors or their assessment</i>			
B. Applicability			
Concern that the definition, assessment or timing of predictors in the model do not match the review question	CONCERN: (low/ high/ unclear)		High
<i>Rationale of applicability rating: They use a previous score and validated that information, but no information of predictors.</i>			
DOMAIN 3: Outcome			
A. Risk of Bias			
<i>Describe the outcome, how it was defined and determined, and the time interval between predictor assessment and outcome determination: Assess the performance of the ADHF/NT-proBNP risk score to predict death in hospitalized patients with AHF, accounting for the competing events VAD implantation and HT.</i>			
		Dev	Val
3.1 Was the outcome determined appropriately?			Y
3.2 Was a pre-specified or standard outcome definition used?			Y
3.3 Were predictors excluded from the outcome definition?			Y

3.4 Was the outcome defined and determined in a similar way for all participants?		Y
3.5 Was the outcome determined without knowledge of predictor information?		Y
3.6 Was the time interval between predictor assessment and outcome determination appropriate?		NI
Risk of bias introduced by the outcome or its determination	RISK: (low/ high/ unclear)	Low
<i>Rationale of bias rating: The outcome was correctly mentioned.</i>		
B. Applicability		
<i>At what time point was the outcome determined: at one time in the future</i>		
<i>If a composite outcome was used, describe the relative frequency/distribution of each contributing outcome:</i>		
Concern that the outcome, its definition, timing or determination do not match the review question	CONCERN: (low/ high/ unclear)	Low
<i>Rationale of applicability rating: The outcome fit with the review question.</i>		
DOMAIN 4: Analysis		
Risk of Bias		
<i>Describe numbers of participants, number of candidate predictors (for DEV only), outcome events and events per candidate predictor (for DEV only): 445 patients. There are no candidate predictors or EPV.</i>		
<i>Describe how the model was developed (predictor selection, optimism, risk groups, model performance): They apply the score in this population. In the entire cohort, the HR for each point of the ADHF/NT -proBNP score for cumulative mortality was 1.24 (95% confidence interval [CI], 1.19–1.29; p o 0.001). For the patients aged r70 years, the HR was 1.28 (95% CI, 1.21– 1.35; p o 0.001). The C-statistic values and the chi-square goodness-of-fit statistics are reported. Predicted and observed 12-month cumulative mortality by quartiles of risk score.</i>		
<i>Describe whether and how the model was validated, either internally (e.g. bootstrapping, cross validation, random split sample) or externally (e.g. temporal validation, geographical validation, different setting, different type of participants): They develop the score by external validation</i>		
<i>Describe the performance measures of the model, e.g. (re)calibration, discrimination, (re)classification, net benefit: The HR for each point of risk score for post-discharge mortality was 1.24 (95% CI, 1.18–1.30; p o 0.001) and was 1.25 (95% CI, 1.17–1.33; p o 0.001) for the patients aged r70 years.</i>		
<i>Describe any participants who were excluded from the analysis: cute coronary syndromes or angina pectoris, recent cardiac surgical or percutaneous procedures, planned coronary revascularization, congenital heart disease, and valvular heart disease regardless of whether surgically corrected.</i>		
<i>Describe missing data on predictors and outcomes as well as methods used for missing data: Missing data, including weight (5.8%), percentage of lymphocytes (12.4%), uric acid (34.6%), and total cholesterol (7.2%), were replaced with the median values of the covariates.</i>		
	Dev	Val

4.1 Were there a reasonable number of participants with the outcome?		Y
4.2 Were continuous and categorical predictors handled appropriately?		Y
4.3 Were all enrolled participants included in the analysis?		Y
4.4 Were participants with missing data handled appropriately?		PY
4.5 Was selection of predictors based on univariable analysis avoided?		Y
4.6 Were complexities in the data (e.g. censoring, competing risks, sampling of controls) accounted for appropriately?		NI
4.7 Were relevant model performance measures evaluated appropriately?		Y
4.8 Were model overfitting and optimism in model performance accounted for?		Y
4.9 Do predictors and their assigned weights in the final model correspond to the results from the reported multivariable analysis?		NI
Risk of bias introduced by the analysis	RISK: (low/ high/ unclear)	Low
<i>Rationale of bias rating: They describe the analysis</i>		

Step 4: Overall assessment

Use the following tables to reach overall judgments about risk of bias and concerns for applicability of the prediction model evaluation (development and/or validation) across all assessed domains.

Reaching an overall judgment about risk of bias of the prediction model evaluation	
Low risk of bias	If all domains were rated low risk of bias. If a <u>prediction model was developed without any external validation</u> , and it was rated as <u>low risk of bias for all domains</u> , consider downgrading to high risk of bias . Such a model can only be considered as low risk of bias, if the development was based on a very large data set and included some form of internal validation.
High risk of bias	If at least one domain is judged to be at high risk of bias .
Unclear risk of bias	If an unclear risk of bias was noted in at least one domain and it was low risk for all other domains.

Complete for each evaluation of a distinct model.

Reaching an overall judgment about applicability of the prediction model evaluation	
Low concerns for applicability	If low concerns for applicability for all domains, the prediction model evaluation is judged to have low concerns for applicability .
High concerns for applicability	If high concerns for applicability for at least one domain, the prediction model evaluation is judged to have high concerns for applicability .
Unclear concerns for applicability	If unclear concerns (but no “high concern”) for applicability for at least one domain, the prediction model evaluation is judged to have unclear concerns for applicability overall.

Overall judgment about risk of bias and applicability of the prediction model evaluation		
Overall judgment of risk of bias	RISK: (low/ high/ unclear)	Unclear

	unclear)	
Summary of sources of potential bias: Unclear		
Overall judgment of applicability	CONCERN: (low/ high/ unclear)	Low
Summary of applicability concerns: Low		

PROBAST = Prediction model Risk Of Bias ASsessment Tool.

33. Shiraishi Y, et al: (2016)

Step 2: Classify the type of prediction model evaluation

Classify the evaluation based on its aim			
Type of prediction study	PROBAST boxes to complete	Tick as appropriate	Definitions for type of prediction model study
Development only	Dev		Prediction model development without external validation. These studies may include internal validation methods such as bootstrapping and cross-validation techniques
Development and validation	Dev and Val		Prediction model development combined with external validation in other participants in the same article
Validation only	Val	X	External validation of existing (previously developed) model in other participants
<i>This table should be completed once for each publication being assessed and for each relevant outcome in your review.</i>			
Publication reference	Shiraishi Y, Kohsaka S, Abe T, Mizuno A, Goda, A. et al. Validation of the Get With The Guideline– Heart Failure risk score in Japanese patients and the potential improvement of its discrimination ability by the inclusion of B-type natriuretic peptide level. American Heart Journal. January 2016. http://dx.doi.org/10.1016/j.ahj.2015.10.008 .		
Models of interest	GWTG-HF score plus BNP.		
Outcome of interest	We aimed to investigate whether the GWTG-HF risk score can be applied to Japanese AHF patients. In addition, we evaluated whether the addition of BNP level into the GWTG-HF risk score improves its predictive ability.		

Step 3: Assess risk of bias and applicability

DOMAIN 1: Participants			
A. Risk of Bias			
Describe the sources of data and criteria for participant selection: The West Tokyo–Heart Failure (WET-HF) registry is a large, ongoing, prospective, multicenter cohort registry designed to collect data on clinical backgrounds and outcomes of patients hospitalized for AHF.			
	De v	Val	
1.1 Were appropriate data sources used, e.g. cohort, RCT or nested case-control study data?		Y	
1.2 Were all inclusions and exclusions of participants appropriate?		N	
Risk of bias introduced by selection of participants	RISK: (low/ high/ unclear)		High
<i>Rationale of bias rating:</i>			
Inclusion and Exclusion criteria were briefly mentioned			
B. Applicability			
Describe included participants, setting and dates: In total, 1,876 patients were admitted for AHF between April 2006 and August 2014. The patients were predominantly men (60.6%), with a mean age of 73.3 ± 13.6 years. Sixty-eight (3.6%) patients died during hospitalization.			
Concern that the included participants and setting do not match the review question	CONCERN: (low/ high/ unclear)		High
<i>Rationale of applicability rating: Information from patients were not correctly mentioned.</i>			
DOMAIN 2: Predictors			
A. Risk of Bias			
List and describe predictors included in the final model, e.g. definition and timing of assessment: SBP, BUN, Sodium, Age, HR, Black race, COPD.			
	De v	Val	
2.1 Were predictors defined and assessed in a similar way for all participants?		Y	
2.2 Were predictor assessments made without knowledge of outcome data?		Y	
2.3 Are all predictors available at the time the model is intended to be used?		Y	
Risk of bias introduced by predictors or their assessment	RISK: (low/ high/ unclear)		Low
<i>Rationale of bias rating: The article has a good selection of variables combining variables known as risk factors in ADHF.</i>			
B. Applicability			
Concern that the definition, assessment or timing of predictors in the model do not match the review question	CONCERN: (low/ high/ unclear)		Low
<i>Rationale of applicability rating: They use a simple group of variables, easy to get for develop one usefull and very applicable model.</i>			

DOMAIN 3: Outcome			
A. Risk of Bias			
Describe the outcome, how it was defined and determined, and the time interval between predictor assessment and outcome determination: We aimed to investigate whether the GWTG-HF risk score can be applied to Japanese AHF patients. In addition, we evaluated whether the addition of BNP level into the GWTG-HF risk score improves its predictive ability.			
		Dev	Val
3.1 Was the outcome determined appropriately?			Y
3.2 Was a pre-specified or standard outcome definition used?			Y
3.3 Were predictors excluded from the outcome definition?			Y
3.4 Was the outcome defined and determined in a similar way for all participants?			Y
3.5 Was the outcome determined without knowledge of predictor information?			Y
3.6 Was the time interval between predictor assessment and outcome determination appropriate?			Y
Risk of bias introduced by the outcome or its determination	RISK: (low/ high/ unclear)		Low
Rationale of bias rating: It was very well defined outcome, with good model			
B. Applicability			
At what time point was the outcome determined: at one time in the future			
If a composite outcome was used, describe the relative frequency/distribution of each contributing outcome:			
Concern that the outcome, its definition, timing or determination do not match the review question	CONCERN: (low/ high/ unclear)		Low
Rationale of applicability rating: It match with the review question.			

DOMAIN 4: Analysis		
Risk of Bias		
Describe numbers of participants, number of candidate predictors (for DEV only), outcome events and events per candidate predictor (for DEV only): ,813 patients who had all of the data required for calculating the score were included in this analysis to validate the GWTG-HF risk score.		
Describe how the model was developed (predictor selection, optimism, risk groups, model performance): We calculated the GWTG-HF risk score to predict all-cause in-hospital mortality for each patient according to the statistical model described in the original article. The variables required for risk score calculation were race, age, systolic blood pressure, heart rate, blood urea nitrogen (BUN) level, sodium concentration, and presence of chronic obstructive pulmonary disease.		
Describe whether and how the model was validated, either internally (e.g. bootstrapping, cross validation, random split sample) or externally (e.g. temporal validation, geographical validation, different setting, different type of participants): External validation of GWTG-HF score in Japanese patients		
Describe the performance measures of the model, e.g. (re)calibration, discrimination, (re)classification, net benefit: The GWTG-HF risk score was a good predictor of in-hospital mortality based on logistic regression analysis (odds ratio per 1-unit increment, 1.122; 95% CI, 1.089-1.156; P b .001). GWTG-HF risk score discrimination was excellent; the area under the receiver operating characteristic curve for the GWTG-HF risk score for in-hospital mortality in this Japanese cohort was 0.763 (95% CI, 0.700-0.826). The calibration of the GWTG-HF risk score was evaluated by comparing observed vs predicted in-hospital mortality. there was good conformance between observed and model-predicted in-hospital mortality across the deciles of predicted risk as determined using the Hosmer-Lemeshow test. However, the calibration plot where the mean model-predicted probability of in-hospital mortality was plotted against the observed in-hospital mortality for the GWTG-HF risk score revealed a deviation from the ideal calibration line with systematic underestimation of risk We conducted an additional analysis to determine whether addition of BNP level would favorably affect the performance of the GWTG-HF risk score. By univariate logistic regression analysis, BNP level at initial presentation was significantly associated with in-hospital mortality (odds ratio, 1.892; 95% CI, 1.462-2.448; P b .001), and the c statistic of BNP level for predicting in-hospital mortality was 0.700 (95% CI, 0.628-0.771). The addition of BNP values to the GWTG-HF risk score significantly changed the score's discrimination ability (c statistic, 0.818; 95% CI, 0.771-0.865). Net reclassification improvement calculations are presented in online. The corresponding calibration plot showed reasonable calibration.		
Describe any participants who were excluded from the analysis: Sixty three (3.3%) patients were excluded because of missing baseline data.		
Describe missing data on predictors and outcomes as well as methods used for missing data: They eliminated patients with missing baseline data		
	Dev	Val
4.1 Were there a reasonable number of participants with the outcome?		Y
4.2 Were continuous and categorical predictors handled appropriately?		Y
4.3 Were all enrolled participants included in the analysis?		Y
4.4 Were participants with missing data handled appropriately?		Y
4.5 Was selection of predictors based on univariable analysis avoided?		Y
4.6 Were complexities in the data (e.g. censoring, competing risks, sampling of controls) accounted for appropriately?		Ni
4.7 Were relevant model performance measures evaluated appropriately?		Y
4.8 Were model overfitting and optimism in model performance accounted for?		Ni
4.9 Do predictors and their assigned weights in the final model correspond to the results from the reported multivariable analysis?		Y

Risk of bias introduced by the analysis	RISK: (low/ unclear) high/		Low
<i>Rationale of bias rating: Describe how they get the score</i>			

Step 4: Overall assessment

Use the following tables to reach overall judgments about risk of bias and concerns for applicability of the prediction model evaluation (development and/or validation) across all assessed domains.

Reaching an overall judgment about risk of bias of the prediction model evaluation	
Low risk of bias	If all domains were rated low risk of bias. If a prediction model was developed without any external validation, and it was rated as <u>low risk of bias for all domains</u> , consider downgrading to high risk of bias . Such a model can only be considered as low risk of bias, if the development was based on a very large data set and included some form of internal validation.
High risk of bias	If at least one domain is judged to be at high risk of bias .
Unclear risk of bias	If an unclear risk of bias was noted in at least one domain and it was low risk for all other domains.

Complete for each evaluation of a distinct model.

Reaching an overall judgment about applicability of the prediction model evaluation	
Low concerns for applicability	If low concerns for applicability for all domains, the prediction model evaluation is judged to have low concerns for applicability .
High concerns for applicability	If high concerns for applicability for at least one domain, the prediction model evaluation is judged to have high concerns for applicability .
Unclear concerns for applicability	If unclear concerns (but no “high concern”) for applicability for at least one domain, the prediction model evaluation is judged to have unclear concerns for applicability overall.

Overall judgment about risk of bias and applicability of the prediction model evaluation		
Overall judgment of risk of bias	RISK: (low/ high/ unclear)	Low
<i>Summary of sources of potential bias: Low</i>		
Overall judgment of applicability	CONCERN: (low/ high/ unclear)	Low
<i>Summary of applicability concerns: Low</i>		

PROBAST = Prediction model Risk Of Bias ASsessment Tool.

34. Shiraishi Y, et al: (2018)

Step 2: Classify the type of prediction model evaluation

Classify the evaluation based on its aim			
Type of prediction study	PROBAST boxes to complete	Tick as appropriate	Definitions for type of prediction model study
Development only	Dev		Prediction model development without external validation. These studies may include internal validation methods such as bootstrapping and cross-validation techniques
Development and validation	Dev and Val		Prediction model development combined with external validation in other participants in the same article
Validation only	Val	X	External validation of existing (previously developed) model in other participants
<i>This table should be completed once for each publication being assessed and for each relevant outcome in your review.</i>			
Publication reference	Shiraishi Y, Kohsaka S, Nagai T, Goda A, Mizuno A, Nagatomo Y, Yasumori Sujino, et al. Validation and Recalibration of Seattle Heart Failure Model in Japanese Acute Heart Failure Patients, Journal of Cardiac Failure.2018. Doi: https://doi.org/10.1016/j.cardfail.2018.07.463		
Models of interest	Seattle Heart Failure Model (SHFM)		
Outcome of interest	Externally validate SHFM in Japanese HF patients with a recent episode of acute decompensation requiring hospital admission		

Step 3: Assess risk of bias and applicability

DOMAIN 1: Participants

A. Risk of Bias			
Describe the sources of data and criteria for participant selection: The participants were included from 2 Japanese AHF registries (WET-HF) and (NaDEF). WET-HF registry collected patients from 2006 to 2016; and NaDEF registry collected from 2013 to 2015. Both of them have patients with the diagnosis of AHF who underwent hospitalization. This is an external validation so they do not show inclusion or exclusion criteria.			
	De v	Val	
1.1 Were appropriate data sources used, e.g. cohort, RCT or nested case-control study data?		Y	
1.2 Were all inclusions and exclusions of participants appropriate?			NI
Risk of bias introduced by selection of participants	RISK: (low/ high/ unclear)		Unclear
Rationale of bias rating:			
Inclusion and exclusion criteria were set in previous articles where they extracted the information.			
B. Applicability			
Describe included participants, setting and dates: The study population consisted of patients with AHF presenting to the investigator's ED in each previous recollection data. The registered patients (mean age 74±13 years) were predominantly men (62%). Overall, 572 (23.2%) patients died 11 during a mean follow-up of 2.1 years.			
Concern that the included participants and setting do not match the review question	CONCERN: (low/ high/ unclear)		Unclear
Rationale of applicability rating: They do not show how was the selection in each previous collection.			

DOMAIN 2: Predictors			
A. Risk of Bias			
List and describe predictors included in the final model, e.g. definition and timing of assessment: Because it is an external validation, the predictors are based on the prior study (SHFM score).			
	De v	Val	
2.1 Were predictors defined and assessed in a similar way for all participants?		Y	
2.2 Were predictor assessments made without knowledge of outcome data?		Y	
2.3 Are all predictors available at the time the model is intended to be used?		Y	
Risk of bias introduced by predictors or their assessment	RISK: (low/ high/ unclear)		Low
Rationale of bias rating: The article use a previous risk model.			
B. Applicability			
Concern that the definition, assessment or timing of predictors in the model do not match the review question	CONCERN: (low/ high/ unclear)		Low
Rationale of applicability rating: They use a group of variables previously demonstrated as a score.			
DOMAIN 3: Outcome			
A. Risk of Bias			

<i>Describe the outcome, how it was defined and determined, and the time interval between predictor assessment and outcome determination: They use a previous developed risk score and now trying to apply it in Japanese population, to identify the best predictors of 1-year mortality.</i>			
		Dev	Val
3.1 Was the outcome determined appropriately?			Y
3.2 Was a pre-specified or standard outcome definition used?			Y
3.3 Were predictors excluded from the outcome definition?			Y
3.4 Was the outcome defined and determined in a similar way for all participants?			Y
3.5 Was the outcome determined without knowledge of predictor information?			Y
3.6 Was the time interval between predictor assessment and outcome determination appropriate?			Y
Risk of bias introduced by the outcome or its determination	RISK: (low/ high/ unclear)		Low
<i>Rationale of bias rating: It was very well defined outcome, with a proven risk model.</i>			
B. Applicability			
<i>At what time point was the outcome determined: at one time in the future</i>			
<i>If a composite outcome was used, describe the relative frequency/distribution of each contributing outcome:</i>			
Concern that the outcome, its definition, timing or determination do not match the review question	CONCERN: (low/ high/ unclear)		Low
<i>Rationale of applicability rating: It was a 100% follow up population.</i>			
DOMAIN 4: Analysis			
Risk of Bias			
<i>Describe numbers of participants, number of candidate predictors (for DEV only), outcome events and events per candidate predictor (for DEV only): 2470 patients in both registries. This article guided its predictors from another study. Candidate predictors are not reported so the EPV cannot be estimated.</i>			
<i>Describe how the model was developed (predictor selection, optimism, risk groups, model performance): They use a previous model</i>			
<i>Describe whether and how the model was validated, either internally (e.g. bootstrapping, cross validation, random split sample) or externally (e.g. temporal validation, geographical validation, different setting, different type of participants): They used a cross validation in several Western countries.</i>			
<i>Describe the performance measures of the model, e.g. (re)calibration, discrimination, (re)classification, net benefit: The calibration and discrimination were assessed using c assessed using the c-statistic and calibration plots. C-statistics of 0.72; and good calibration (we do not have the Calibration Plots) in the entire cohort. However poor calibration in patients with HFrEF and HFpEF</i>			
<i>Describe any participants who were excluded from the analysis: They do not specify exclusion criteria.</i>			
<i>Describe missing data on predictors and outcomes as well as methods used for missing data: they have a few missing data percentage (3%) exception of total cholesterol (9.9%), lymphocytes (6.4%), and left ventricular ejection fraction (LVEF) (3.3%),</i>			
		Dev	Val
4.1 Were there a reasonable number of participants with the outcome?			Y

4.2 Were continuous and categorical predictors handled appropriately?		NI
4.3 Were all enrolled participants included in the analysis?		Y
4.4 Were participants with missing data handled appropriately?		Y
4.5 Was selection of predictors based on univariable analysis avoided?		NI
4.6 Were complexities in the data (e.g. censoring, competing risks, sampling of controls) accounted for appropriately?		NI
4.7 Were relevant model performance measures evaluated appropriately?		Y
4.8 Were model overfitting and optimism in model performance accounted for?		Y
4.9 Do predictors and their assigned weights in the final model correspond to the results from the reported multivariable analysis?		Y
Risk of bias introduced by the analysis	RISK: (low/ high/ unclear)	Unclear
<i>Rationale of bias rating: Do not describe in the present study some aspects like predictors selection. They presented values of c statistics but do not showed calibration plots</i>		

Step 4: Overall assessment

Use the following tables to reach overall judgments about risk of bias and concerns for applicability of the prediction model evaluation (development and/or validation) across all assessed domains.

Reaching an overall judgment about risk of bias of the prediction model evaluation	
Low risk of bias	If all domains were rated low risk of bias. If a prediction model was developed without any external validation, and it was rated as low risk of bias for all domains , consider downgrading to high risk of bias . Such a model can only be considered as low risk of bias, if the development was based on a very large data set and included some form of internal validation.
High risk of bias	If at least one domain is judged to be at high risk of bias .
Unclear risk of bias	If an unclear risk of bias was noted in at least one domain and it was low risk for all other domains.

Complete for each evaluation of a distinct model.

Reaching an overall judgment about applicability of the prediction model evaluation	
Low concerns for applicability	If low concerns for applicability for all domains, the prediction model evaluation is judged to have low concerns for applicability .
High concerns for applicability	If high concerns for applicability for at least one domain, the prediction model evaluation is judged to have high concerns for applicability .
Unclear concerns for applicability	If unclear concerns (but no "high concern") for applicability for at least one domain, the prediction model evaluation is judged to have unclear concerns for applicability overall.

Overall judgment about risk of bias and applicability of the prediction model evaluation		
Overall judgment of risk of bias	RISK: (low/ high/ unclear)	Unclear

Summary of sources of potential bias: Unclear		
Overall judgment of applicability	CONCERN: (low/ high/ unclear)	Unclear
Summary of applicability concerns: Unclear		

PROBAST = Prediction model Risk Of Bias ASsessment Tool.

35. Spinar J, et al:

Step 2: Classify the type of prediction model evaluation

Classify the evaluation based on its aim			
Type of prediction study	PROBAST boxes to complete	Tick as appropriate	Definitions for type of prediction model study
Development only	Dev		Prediction model development without external validation. These studies may include internal validation methods such as bootstrapping and cross-validation techniques
Development and validation	Dev and Val	X	Prediction model development combined with external validation in other participants in the same article
Validation only	Val		External validation of existing (previously developed) model in other participants
<i>This table should be completed once for each publication being assessed and for each relevant outcome in your review.</i>			
Publication reference	Spinar J, Jarkovsky J, Spinarova L, Mebazaa A, Gayat E, Vitovec J, et al. AHEAD score — Long-term risk classification in acute heart failure. International Journal of Cardiology.2016. 202:21–26		
Models of interest	AHEAD score		
Outcome of interest	Main outcome was one year all-cause mortality		

Step 3: Assess risk of bias and applicability

DOMAIN 1: Participants
A. Risk of Bias
Describe the sources of data and criteria for participant selection: The Acute Heart Failure Database (AHEAD) Network registry includes consecutive patients from ten centres with 24-h Catheter Laboratory services and centralised care for patients with acute coronary syndromes (described previously as “AHEAD Main”) and from five regional centres without a Catheter Laboratory from a region of approximately 3 million inhabitants of the Czech Republic. The inclusion criteria for the database were in conformance to the European guidelines for AHF set in 2005 and were as follows: signs and symptoms of AHF; confirmed left-ventricular dysfunction (systolic or diastolic); and/or positive response to therapy [15]. Exclusion criteria were known or newly diagnosed advanced stage of malignancy, and disagreement with participation in the registry. The AHEAD Network registry includes 7318 hospitalisations for AHF with 6242 primo hospitalisations for AHF; the dataset

of AHF primo hospitalisations with available key variables (N=5846) was adopted for the AHEAD score analysis — derivation cohort. The validation AHF dataset of the GREAT registry consists of nine cohorts from Italy (N = 1828), Spain (N = 1631), France (N = 696), Argentina (N = 675), Finland (N = 584), Switzerland (N = 370), USA (N = 209), Tunisia (N = 186), and Austria (N = 136). One-year all cause mortality data were available for all patients.			
		De v	Val
1.1 Were appropriate data sources used, e.g. cohort, RCT or nested case-control study data?		Y	Y
1.2 Were all inclusions and exclusions of participants appropriate?		Y	Y
Risk of bias introduced by selection of participants	RISK: (low/ high/ unclear)	Low	Low
<i>Rationale of bias rating: Applied correct criteria for inclusion of participants</i>			
B. Applicability			
<i>Describe included participants, setting and dates:</i> The baseline characteristics of the AHEAD registry (derivation cohort, n = 5846) and GREAT dataset (validation cohort, n = 6315). Compared with the GREAT cohort, patients in the AHEAD cohort were slightly younger (median age, 74 vs. 77 years), had a lower rate of anaemia (38.2% vs. 50.2%) and AF (30.7% vs. 39.0%), and had a higher rate of elevated creatinine (N130 µmol/l) (30.1% vs. 22.0%), hypertension (75.0% vs. 70%) and diabetes mellitus (44.0% vs. 38.0%). Medications upon hospital admission and in surviving patients at discharge were similar. The 30 day and 1 year mortality rates of the AHEAD vs. the GREAT cohort were 13.7% vs. 14.4 and 32.3% vs. 31.6%, respectively. In the AHEAD cohort, the 3 and 5 year mortality rates were 48.4% and 60.3%, respectively.			
Concern that the included participants and setting do not match the review question	CONCERN: (low/ high/ unclear)	Low	Low
<i>Rationale of applicability rating: The characteristics of the study population are consistent with the criteria of the research question.</i>			

DOMAIN 2: Predictors			
A. Risk of Bias			
<i>List and describe predictors included in the final model, e.g. definition and timing of assessment:</i> (A) atrial fibrillation, anaemia (H — haemoglobin level less than 130 g/l for men and 120 g/l for women), (E) elderly (age ≥70 years), (A) abnormal renal parameters (creatinine ≥130 µmol/l) and (D) diabetes mellitus.			
		De v	Val
2.1 Were predictors defined and assessed in a similar way for all participants?		Y	Y
2.2 Were predictor assessments made without knowledge of outcome data?		Y	Y
2.3 Are all predictors available at the time the model is intended to be used?		Y	Y
Risk of bias introduced by predictors or their assessment	RISK: (low/ high/ unclear)	Low	Low
<i>Rationale of bias rating: They adequately describe the way of choosing final predictors.</i>			
B. Applicability			
Concern that the definition, assessment or timing of predictors in the model do not match the review question	CONCERN: (low/ high/ unclear)	Low	Low
<i>Rationale of applicability rating: The evaluation time of the predictive variables and the applicability of these are consistent with the research question of the recent systematic review.</i>			
DOMAIN 3: Outcome			

A. Risk of Bias			
Describe the outcome, how it was defined and determined, and the time interval between predictor assessment and outcome determination: The model seeks to predict mortality from any cause for one year in patients with AHF.			
		Dev	Val
3.1 Was the outcome determined appropriately?		PY	PY
3.2 Was a pre-specified or standard outcome definition used?		Y	Y
3.3 Were predictors excluded from the outcome definition?		Y	Y
3.4 Was the outcome defined and determined in a similar way for all participants?		Y	Y
3.5 Was the outcome determined without knowledge of predictor information?		Y	Y
3.6 Was the time interval between predictor assessment and outcome determination appropriate?		Y	Y
Risk of bias introduced by the outcome or its determination	RISK: (low/ high/ unclear)	Low	Low
Rationale of bias rating: Adequate description of the outcomes and the manner in which they were collected.			
B. Applicability			
At what time point was the outcome determined: It was determined during the follow-up in the future, during the time of hospitalization. If a composite outcome was used, describe the relative frequency/distribution of each contributing outcome:			
Concern that the outcome, its definition, timing or determination do not match the review question	CONCERN: (low/ high/ unclear)	Low	Low
Rationale of applicability rating: The time in which the outcome is determined is consistent with the proposed research question.			
DOMAIN 4: Analysis			
Risk of Bias			
Describe numbers of participants, number of candidate predictors (for DEV only), outcome events and events per candidate predictor (for DEV only): The baseline characteristics of the AHEAD registry (derivation cohort, n = 5846) and GREAT dataset (validation cohort, n = 6315). Compared with the GREAT cohort, patients in the AHEAD cohort were slightly younger (median age, 74 vs. 77 years), had a lower rate of anaemia (38.2% vs. 50.2%) and AF (30.7% vs. 39.0%), and had a higher rate of elevated creatinine (N130 µmol/l) (30.1% vs. 22.0%), hypertension (75.0% vs. 70%) and diabetes mellitus (44.0% vs. 38.0%). Medications upon hospital admission and in surviving patients at discharge were similar. The 30 day and 1 year mortality rates of the AHEAD vs. the GREAT cohort were 13.7% vs. 14.4 and 32.3% vs. 31.6%, respectively. In the AHEAD cohort, the 3 and 5 year mortality rates were 48.4% and 60.3%, respectively. They don't say it explicitly, but according to their table of variables it is likely that the model initially had 13 candidate variables for predictors. The 30 day and 1 year mortality rates of the AHEAD vs. the GREAT cohort were 13.7% vs. 14.4 and 32.3% vs. 31.6%, respectively.			
Describe how the model was developed (predictor selection, optimism, risk groups, model performance): Standard descriptive statistics were applied in the analysis; absolute and relative frequencies for categorical variables and median supplemented by 5th–95th percentiles for continuous variables. The predictive power of patients' characteristics for the survival of patients was computed using the Cox proportional hazards model and described by hazard ratios and their 95% confidence intervals. The multivariate Cox proportional hazards model, combining expert selection of predictors with a forward stepwise selection algorithm, was used for the definition of the variables included in the computation of the AHEAD score; the selection of variables for the alternative statistical model was based on the forward stepwise selection algorithm alone. Both scores were			

<p>computed as the sum of the risk factors identified by multivariate analysis. The survival according to scores was visualised by Kaplan–Meier methodology and compared using the log-rank test. Survival rates at given time points were computed using Kaplan–Meier methodology. For the description of the predictive power of scores at given time points, the time dependent ROC methodology was adopted. A level of $\alpha = 0.05$ was used as a boundary for statistical significance in all analyses. Statistical analyses were performed using SPSS 22.0.0.1 statistical package (IBM Corporation 2013) and R version 3.1.1 with timeROC library (The R Foundation for Statistical Computing, 2014).</p>			
<p><i>Describe whether and how the model was validated, either internally (e.g. bootstrapping, cross validation, random split sample) or externally (e.g. temporal validation, geographical validation, different setting, different type of participants):</i> Derivation and validation cohorts were used for internal and external validation respectively.</p>			
<p><i>Describe the performance measures of the model, e.g. (re)calibration, discrimination, (re)classification, net benefit:</i> Patients were divided according to the number of points into six groups for the AHEAD score and seven groups for the statistical model (patients with 6–8 points were included in the ≥ 6 group because of the small numbers of patients with 7 or 8 points). Using time dependent receiver operating characteristics (ROC) analysis, the area under the curve (AUC) was calculated for comparison of the predictive value between the AHEAD score and maximal statistical model at selected time points. In both models, there was a clear trend for better predictive value with longer time from onset of AHF (at 12 months, the AUCs for the AHEAD score and the statistical model were 0.639 and 0.655, respectively, whereas those at 5 years were 0.701 and 0.722, respectively). The statistical model showed a slightly better predictive value than the AHEAD score, and during the first 5 years, the predictive values of both models were comparable. The predictive value of the AHEAD score during the first 12 months after AHF was externally validated on the GREAT dataset. The AUC of time dependent c statistics for the GREAT dataset was 0.631.</p>			
<p><i>Describe any participants who were excluded from the analysis:</i> Exclusion criteria were known or newly diagnosed advanced stage of malignancy, and disagreement with participation in the registry. The characteristics of excluded patients are not mentioned.</p>			
<p><i>Describe missing data on predictors and outcomes as well as methods used for missing data:</i> The handling of missing data is not mentioned.</p>			
		Dev	Val
4.1 Were there a reasonable number of participants with the outcome?		Y	Y
4.2 Were continuous and categorical predictors handled appropriately?		Y	Y
4.3 Were all enrolled participants included in the analysis?		PY	PY
4.4 Were participants with missing data handled appropriately?		NI	NI
4.5 Was selection of predictors based on univariable analysis avoided?		Y	Y
4.6 Were complexities in the data (e.g. censoring, competing risks, sampling of controls) accounted for appropriately?		PY	PY
4.7 Were relevant model performance measures evaluated appropriately?		PY	PY
4.8 Were model overfitting and optimism in model performance accounted for?		NI	NI
4.9 Do predictors and their assigned weights in the final model correspond to the results from the reported multivariable analysis?		Y	Y
Risk of bias introduced by the analysis	RISK: (low/ high/ unclear)	Low	Low
<p><i>Rationale of bias rating:</i> The article doesn't evaluate the calibration within the performance, however, it adequately describes the discrimination and the statistical analysis performed.</p>			

Step 4: Overall assessment

Use the following tables to reach overall judgments about risk of bias and concerns for applicability of the prediction model evaluation (development and/or validation) across all assessed domains.

Reaching an overall judgment about risk of bias of the prediction model evaluation	
Low risk of bias	If all domains were rated low risk of bias. If a prediction model was developed without any external validation, and it was rated as <u>low risk of bias for all domains</u> , consider downgrading to high risk of bias . Such a model can only be considered as low risk of bias, if the development was based on a very large data set and included some form of internal validation.
High risk of bias	If at least one domain is judged to be at high risk of bias .
Unclear risk of bias	If an unclear risk of bias was noted in at least one domain and it was low risk for all other domains.

Complete for each evaluation of a distinct model.

Reaching an overall judgment about applicability of the prediction model evaluation	
Low concerns for applicability	If low concerns for applicability for all domains, the prediction model evaluation is judged to have low concerns for applicability .
High concerns for applicability	If high concerns for applicability for at least one domain, the prediction model evaluation is judged to have high concerns for applicability .
Unclear concerns for applicability	If unclear concerns (but no “high concern”) for applicability for at least one domain, the prediction model evaluation is judged to have unclear concerns for applicability overall.

Overall judgment about risk of bias and applicability of the prediction model evaluation		
Overall judgment of risk of bias	RISK: (low/ high/ unclear)	
<i>Summary of sources of potential bias: Low</i>		
Overall judgment of applicability	CONCERN: (low/ high/ unclear)	
<i>Summary of applicability concerns: Low</i>		

36. Thorsvaldsen T, et al.

Step 2: Classify the type of prediction model evaluation

Classify the evaluation based on its aim

Type of prediction study	PROBAST boxes to complete	Tick as appropriate	Definitions for type of prediction model study
Development only	Dev		Prediction model development without external validation. These studies may include internal validation methods such as bootstrapping and cross-validation techniques
Development and validation	Dev and Val	X	Prediction model development combined with external validation in other participants in the same article
Validation only	Val		External validation of existing (previously developed) model in other participants
<i>This table should be completed once for each publication being assessed and for each relevant outcome in your review.</i>			
Publication reference	Thorsvaldsen T. Claggett B. Shah A. Cheng S. Agarwal S. Predicting Risk in Patients Hospitalized for Acute Decompensated Heart Failure and Preserved Ejection Fraction. The Atherosclerosis Risk in Communities Study Heart Failure Community Surveillance. Circ Heart Fail. 2017;10:e003992. DOI: 10.1161/CIRCHEARTFAILURE.117.003992		
Models of interest	Predicting Risk in Patients Hospitalized for Acute Decompensated Heart Failure and Preserved Ejection Fraction.		
Outcome of interest	Identify predictors of mortality and to create a risk prediction model in patients with HFpEF hospitalized for ADHF.		

Step 3: Assess risk of bias and applicability

DOMAIN 1: Participants			
A. Risk of Bias			
Describe the sources of data and criteria for participant selection: the Heart Failure Community Surveillance component of the ARIC study enumerates and validates HF hospitalizations from 21 hospitals from 4 United States communities (Forsyth County, NC; Jackson, MS; suburbs of Minneapolis, MN; and Washington County, MD). Methods of event ascertainment and classification have been described previously. ¹⁰ A stratified random sample of eligible hospitalizations for HF is selected based on a HF-related International Classification of Disease, Ninth Revision, Clinical Modification codes in any position; age ≥55 years at the time of hospital discharge, and home address within the ARIC communities. Abstracted hospitalizations are classified by physician review or computer algorithm as definite or possible ADHF, chronic stable HF, and HF unlikely or unclassifiable.			
		Dev	Val
1.1 Were appropriate data sources used, e.g. cohort, RCT or nested case-control study data?		Y	Y
1.2 Were all inclusions and exclusions of participants appropriate?		Y	Y
Risk of bias introduced by selection of participants		RISK: (low/ high/ unclear)	
		Low	Low
<i>Rationale of bias rating:</i>			
Inclusion criteria was briefly mentioned, but exclusion criteria was very well defined.			
B. Applicability			

<i>Describe included participants, setting and dates: The study population consisted of patients with AHF.</i>			
Concern that the included participants and setting do not match the review question	CONCERN: (low/ high/ unclear)	Low	Low
<i>Rationale of applicability rating: Information from patients were correctly mentioned.</i>			

DOMAIN 2: Predictors			
A. Risk of Bias			
<i>List and describe predictors included in the final model, e.g. definition and timing of assessment: age, SBP, BUN, Sodium, cerebrovascular disease, COPD plus hb, HR, Natriuretic peptides, hypoxia, white race, BMI.</i>			
		Dev	Val
2.1 Were predictors defined and assessed in a similar way for all participants?		Y	Y
2.2 Were predictor assessments made without knowledge of outcome data?		Y	y
2.3 Are all predictors available at the time the model is intended to be used?		Y	y
Risk of bias introduced by predictors or their assessment	RISK: (low/ high/ unclear)	Low	Low
<i>Rationale of bias rating: The article has a good selection of variables combining variables known as risk factors in ADHF.</i>			
B. Applicability			
Concern that the definition, assessment or timing of predictors in the model do not match the review question	CONCERN: (low/ high/ unclear)	Low	Low
<i>Rationale of applicability rating: They use a simple group of variables, easy to get for develop one usefull and very applicable model.</i>			
DOMAIN 3: Outcome			
A. Risk of Bias			
<i>Describe the outcome, how it was defined and determined, and the time interval between predictor assessment and outcome determination: Identify predictors of mortality and to create a risk prediction model in patients with HFpEF hospitalized for ADHF. They use this predictor to predict 28 days and 1-year risk mortality.</i>			
		Dev	Val
3.1 Was the outcome determined appropriately?		Y	Y
3.2 Was a pre-specified or standard outcome definition used?		Y	Y
3.3 Were predictors excluded from the outcome definition?		Y	Y
3.4 Was the outcome defined and determined in a similar way for all participants?		Y	Y
3.5 Was the outcome determined without knowledge of predictor information?		Y	Y
3.6 Was the time interval between predictor assessment and outcome determination appropriate?		Y	Y
Risk of bias introduced by the outcome or its determination	RISK: (low/ high/ unclear)	Low	Low
<i>Rationale of bias rating: It was very well defined outcome, with good model</i>			

B. Applicability			
<i>At what time point was the outcome determined: at one time in the future</i>			
<i>If a composite outcome was used, describe the relative frequency/distribution of each contributing outcome:</i>			
Concern that the outcome, its definition, timing or determination do not match the review question	CONCERN: (low/ high/ unclear)	Low	Low
<i>Rationale of applicability rating: It was correctly defined the outcome and the time; and It match the review question</i>			
DOMAIN 4: Analysis			
Risk of Bias			
<i>Describe numbers of participants, number of candidate predictors (for DEV only), outcome events and events per candidate predictor (for DEV only): 2673 patients, 1852 derivation sample and 821 in validation sample.</i>			
<i>Describe how the model was developed (predictor selection, optimism, risk groups, model performance): From the identified predictors, separate risk scores for 28-day and 1-year mortalities were created. The probability of dying within 28 days and 1 year was estimated by summing the points assigned for each value of the predictors. For 28-day mortality, the median risk score was 44 with a range of 9 to 84. For 1-year mortality, the median risk score was 29, range 8 to 59.</i>			
<i>Describe whether and how the model was validated, either internally (e.g. bootstrapping, cross validation, random split sample) or externally (e.g. temporal validation, geographical validation, different setting, different type of participants): They had a derivation cohort were developed the score, and a validation cohort for internal validation.</i>			
<i>Describe the performance measures of the model, e.g. (re)calibration, discrimination, (re)classification, net benefit: Mortality in the validation sample was 8% at 28 days and 31% at 1 year, the corresponding mean predicted numbers using the risk score were 7% and 29%. Risk score discrimination in the validation sample was only slightly weaker than in the derivation sample (AUC 0.73 and 0.71 for 28-day and 1-year mortalities, respectively). Calibration was acceptable as shown with Hosmer–Lemeshow statistics (risk score P value 0.51 and 0.29 for 28-day and 1-year mortalities) and the plots of predicted versus observed mortality</i>			
<i>Describe any participants who were excluded from the analysis: hospitalizations with EF <50%, EF ≥50% with prior EF <50%, and EF missing. We further excluded hos- pitalizations of ADHF developed during hospitalization (rather than at time of admission), with unknown status at follow- up, and race other than black or white</i>			
<i>Describe missing data on predictors and outcomes as well as methods used for missing data: For variables with >5% missing, was performed simple imputation using the sample mean for missing values.</i>			
	Dev	Val	
4.1 Were there a reasonable number of participants with the outcome?	Y	Y	
4.2 Were continuous and categorical predictors handled appropriately?	Y	Y	
4.3 Were all enrolled participants included in the analysis?	Y	Y	
4.4 Were participants with missing data handled appropriately?	y	Y	
4.5 Was selection of predictors based on univariable analysis avoided?	Y	Y	
4.6 Were complexities in the data (e.g. censoring, competing risks, sampling of controls)	Ni	Ni	

accounted for appropriately?		
4.7 Were relevant model performance measures evaluated appropriately?	Y	Y
4.8 Were model overfitting and optimism in model performance accounted for?	Y	Y
4.9 Do predictors and their assigned weights in the final model correspond to the results from the reported multivariable analysis?	Y	Y
Risk of bias introduced by the analysis	RISK: (low/ high/ unclear)	Low Low
<i>Rationale of bias rating: Describe how they get the score</i>		

Step 4: Overall assessment

Use the following tables to reach overall judgments about risk of bias and concerns for applicability of the prediction model evaluation (development and/or validation) across all assessed domains.

Reaching an overall judgment about risk of bias of the prediction model evaluation	
Low risk of bias	If all domains were rated low risk of bias. If a <u>prediction model</u> was developed without any external validation, and it was rated as <u>low risk of bias for all domains</u> , consider downgrading to high risk of bias . Such a model can only be considered as low risk of bias, if the development was based on a very large data set and included some form of internal validation.
High risk of bias	If at least one domain is judged to be at high risk of bias .
Unclear risk of bias	If an unclear risk of bias was noted in at least one domain and it was low risk for all other domains.

Complete for each evaluation of a distinct model.

Reaching an overall judgment about applicability of the prediction model evaluation	
Low concerns for applicability	If low concerns for applicability for all domains, the prediction model evaluation is judged to have low concerns for applicability .
High concerns for applicability	If high concerns for applicability for at least one domain, the prediction model evaluation is judged to have high concerns for applicability .
Unclear concerns for applicability	If unclear concerns (but no “high concern”) for applicability for at least one domain, the prediction model evaluation is judged to have unclear concerns for applicability overall.

Overall judgment about risk of bias and applicability of the prediction model evaluation		
Overall judgment of risk of bias	RISK: (low/ high/ unclear)	Low
<i>Summary of sources of potential bias: Low</i>		
Overall judgment of applicability	CONCERN: (low/ high/	Low

	unclear)	
Summary of applicability concerns: Low		

37. Tomcikova D, et al:

Step 2: Classify the type of prediction model evaluation

Classify the evaluation based on its aim			
Type of prediction study	PROBAST boxes to complete	Tick as appropriate	Definitions for type of prediction model study
Development only	Dev		Prediction model development without external validation. These studies may include internal validation methods such as bootstrapping and cross-validation techniques
Development and validation	Dev and Val	X	Prediction model development combined with external validation in other participants in the same article
Validation only	Val		External validation of existing (previously developed) model in other participants
<i>This table should be completed once for each publication being assessed and for each relevant outcome in your review.</i>			
Publication reference	Tomcikova D, Felsoci M, Spinar J, Miklik R, Mikusova T, Vitovec J, et al. Risk of in-hospital mortality identified according to the typology of patients with acute heart failure: Classification tree analysis on data from the Acute Heart Failure Database–Main registry. Journal of Critical Care.2013.28:250–258		
Models of interest	Risk of in-hospital mortality identified according to the typology of patients with acute heart failure: Classification tree analysis on data from the Acute Heart Failure Database–Main registry		
Outcome of interest	Recognize the strongest parameters regarding in-hospital mortality		

Step 3: Assess risk of bias and applicability

DOMAIN 1: Participants
A. Risk of Bias
<i>Describe the sources of data and criteria for participant selection: The AHEAD registry consists of 2 independent parts. The AHEAD-Main registry includes all consecutive patients treated in 7 centers with a 24-hour catheterization laboratory service. The AHEAD-Network registry comprises 5 regional hospitals without a catheterization laboratory service. The project AHEAD adopted sophisticated technological and managerial rules fully implemented according to ISO standards (CSN ISO/IEC 27001:2006 on the information security management), which also covers quality control/ quality assurance mechanisms. The database records are updated continuously, directly in routine practice of the participating centers, and quarterly, the records should be controlled by local data managers against health care documentation. Second level of controls is realized as a part of validation procedures of the centralized data management team and includes analytical control of consistency of records in time, completeness of the records, and correctness of quantitative values. The current analysis was performed only on the data from the AHEAD-Main registry, which data originate from cardiology departments. The inclusion criteria, the baseline characteristic of the study population, and inhospital mortality</i>

were published elsewhere. Briefly, the patients' population involves all clinical presentation syndromes of AHF according to European guidelines for AHF: (1) acute decompensated HF (ADHF) (with signs and symptoms of AHF, which are mild and do not fulfill criteria for CS, pulmonary edema, or hypertensive crisis); (2) hypertensive AHF (symptoms of AHF are accompanied by high blood pressure (BP) on admission and relatively preserved left ventricular function with a chest radiograph compatible with acute pulmonary edema); (3) pulmonary edema (accompanied by severe respiratory distress, with crackles over the lungs and orthopnea with O2 saturation usually $\geq 90\%$ prior treatment); (4) CS (defined as evidence of tissue hypoperfusion induced by HF after correction of preload, mostly with SBP ≥ 90 mm Hg ongoing for at least 30 minutes); (5) high output failure (characterized by high cardiac output, usually with high heart rate often caused by arrhythmias, thyrotoxicosis crisis, and anemia); and (6) right HF (characterized by low output syndrome with increased jugular venous pressure, increased liver size and hypotension. The study protocol complied with the Declaration of Helsinki, and an informed consent was obtained from all the participants in the registry and local ethics committees of the involved centers approved the database protocol.

		De v	Val
1.1 Were appropriate data sources used, e.g. cohort, RCT or nested case-control study data?		Y	Y
1.2 Were all inclusions and exclusions of participants appropriate?		Y	Y
Risk of bias introduced by selection of participants	RISK: (low/ high/ unclear)	Low	Low
<i>Rationale of bias rating: Applied correct criteria for inclusion of participants</i>			
B. Applicability			
<i>Describe included participants, setting and dates:</i> The current analysis is based on the data of 4153 patients' records between September 2006 and October 2009. The analysis involves only the first hospitalization due to AHF for each patient during the study period. The analyzed cohort also included patients hospitalized for AHF in a pilot study of the biggest center—University Hospital Brno, which started the enrollment earlier, in December 2004. Of 4153 patients, 12.7% died during hospitalization. According to the classification of HF syndromes, inhospital mortality was 2.5% in ADHF, 2.2% in hypertensive AHF, 7.1% in pulmonary edema, 62.7% in CS, 6.1% in high output failure, and 16.7% in right HF. From the whole sample of 4153 patients, 600 of them (14.4%) had signs of CS (acute coronary syndrome [ACS] was confirmed for 371 of them).			
Concern that the included participants and setting do not match the review question	CONCERN: (low/ high/ unclear)	Low	Low
<i>Rationale of applicability rating: The characteristics of the study population are consistent with the criteria of the research question.</i>			

DOMAIN 2: Predictors

A. Risk of Bias

List and describe predictors included in the final model, e.g. definition and timing of assessment: Presence of cardiogenic shock (CS), the level of creatinine, and the systolic blood pressure (SBP). A total of 27 clinical and laboratory variables were tested as potential predictors of in-hospital mortality; the variables were selected from 165 parameters recorded in the registry database. The selection of tested variables was performed by 2 experienced cardiologists and based on previous published result, the tree was verified on the validation cohort.

		De v	Val
2.1 Were predictors defined and assessed in a similar way for all participants?		Y	Y
2.2 Were predictor assessments made without knowledge of outcome data?		Y	Y
2.3 Are all predictors available at the time the model is intended to be used?		Y	Y
Risk of bias introduced by predictors or their assessment	RISK: (low/ high/	Low	Low

	unclear)		
Rationale of bias rating: describes a reliable method of selecting final variables.			
B. Applicability			
Concern that the definition, assessment or timing of predictors in the model do not match the review question	CONCERN: (low/ high/ unclear)	Low	Low
Rationale of applicability rating: The evaluation time of the predictive variables and the applicability of these are consistent with the research question of the recent systematic review.			
DOMAIN 3: Outcome			
A. Risk of Bias			
Describe the outcome, how it was defined and determined, and the time interval between predictor assessment and outcome determination: The outcome (dependent, end point) variable was in-hospital mortality coded in binary categories: 0, discharge; 1, death during hospitalization. The principle of the classification tree method (CART) is the subsequent binary splitting of patients into groups with the highest contrast in the evaluated outcome.			
		Dev	Val
3.1 Was the outcome determined appropriately?		Y	Y
3.2 Was a pre-specified or standard outcome definition used?		Y	Y
3.3 Were predictors excluded from the outcome definition?		Y	Y
3.4 Was the outcome defined and determined in a similar way for all participants?		Y	Y
3.5 Was the outcome determined without knowledge of predictor information?		Y	Y
3.6 Was the time interval between predictor assessment and outcome determination appropriate?		Y	Y
Risk of bias introduced by the outcome or its determination	RISK: (low/ high/ unclear)	Low	Low
Rationale of bias rating: Adequate description of the outcomes and the manner in which they were collected.			
B. Applicability			
At what time point was the outcome determined: It was determined during the follow-up in the future, during the time of hospitalization.			
If a composite outcome was used, describe the relative frequency/distribution of each contributing outcome:			
Concern that the outcome, its definition, timing or determination do not match the review question	CONCERN: (low/ high/ unclear)	Low	Low
Rationale of applicability rating: The time in which the outcome is determined is consistent with the proposed research question.			
DOMAIN 4: Analysis			
Risk of Bias			
Describe numbers of participants, number of candidate predictors (for DEV only), outcome events and events per candidate predictor (for DEV only): It has a total of 4153 participants. They doesn't provide an exact number of events in the population analyzed. A total of 27 clinical and laboratory variables were tested as potential predictors of in-hospital mortality; the variables were selected from 165 parameters recorded in the			

registry database.

Describe how the model was developed (predictor selection, optimism, risk groups, model performance): The CART analysis was applied as nonparametric method, which does not assume specific distribution of the used variables. The outcome (dependent, end point) variable was in-hospital mortality coded in binary categories: 0, discharge; 1, death during hospitalization. The principle of the classification tree method (CART) is the subsequent binary splitting of patients into groups with the highest contrast in the evaluated outcome. In each level of the tree, the variable with the strongest relationship to the end point is selected, and the optimal cut-off value is used. The selected cut-off value splits the cohort into most homogenous groups given the maximum contrast between these 2 groups with regard to the outcome variable. The same algorithm is run consequently on the groups newly formed in this way, and the calculation continues until the stopping rule is reached, that is, the minimum of 50 cases split into groups in our case. If further split does not increase homogeneity with regard to the outcome, the algorithm stops. Our tree was grown using Gini index as the measure of impurity. The 10 fold cross validation was applied to select the final tree, minimizing the misclassification error. No imputation technique of the missing values and no surrogate variables were used while tree growing. Thus, we exclude 223 (5.4%) of 4153 records from the analysis due to missing information regarding any of the predictive variables used in the classification tree. The available data set from AHEAD Main registry was split into the derivation (2/3 of records; n = 2543) and the validation cohort (1/3 of records; n = 1387) by simple random sampling. The tree model was calibrated on the derivation cohort. The prior probabilities of the outcome variable were estimated for the classification tree. A total of 27 clinical and laboratory variables were tested as potential predictors of in-hospital mortality; the variables were selected from 165 parameters recorded in the registry database. The selection of tested variables was performed by 2 experienced cardiologists and based on previous published results. Subsequently, the tree was verified on the validation cohort. Because of the relatively low predictive ability of CART reported previously, we decided for further assessment of the model: (1) the discrimination ability of the model between risk groups was further assessed by odds ratios (ORs), (2) the ability of the model to predict the outcome (survivor or deceased during hospitalization) was also evaluated by the area under the receiver operating characteristics (ROC) curve (c-statistics), and (3) the performance of the model was evaluated by Brier's score to assess the agreement between model's probability prediction and the real outcome; the lower Brier's score, the better the performance of the model. For computing area under ROC and Brier's score, we used the binary categories of the outcome variable assigned to the final groups in tree. As a final step of additional model assessment, the predicted risk groups were related to the other clinical indicators of the disease severity: the etiology and parameters indicating adverse event during hospitalization. For testing statistical differences among groups, the Mann-Whitney U test and the Kruskal-Wallis test for continuous variables and the χ^2 or Fisher exact test for categorical variables were used. In addition, the 1-way analysis of variance and the χ^2 trend test were used for trend testing. In the analyses, the following software was used: SPSS 20.0.1 (IBM Corporation, 2010, New York, NY), STATISTICA (StatSoft, Inc, 2010, Tulsa, OK) and R (package tree) for classification tree analysis.

Describe whether and how the model was validated, either internally (e.g. bootstrapping, cross validation, random split sample) or externally (e.g. temporal validation, geographical validation, different setting, different type of participants): Cross validation was used for internal validation of the model, and for external validation a different population was used for the derivation cohort.

Describe the performance measures of the model, e.g. (re)calibration, discrimination, (re)classification, net benefit: The presence of CS was the strongest predictor of in-hospital mortality. Then, for the creatinine blood level at admission, the best split-level was 140 $\mu\text{mol/L}$ in cases with CS and 155 $\mu\text{mol/L}$ in patients without CS. The SBP at admission served as the last variable significantly contributing to the prediction in patients without CS, with the cut-off value 103 mm Hg. The risk of in-hospital mortality ranged from 2.8% in patients without CS and low levels of creatinine less than 155 $\mu\text{mol/L}$, to 66.2% in patients with CS with creatinine levels greater than 140 $\mu\text{mol/L}$. The tree was subsequently validated. For the category of discharged patients (in-hospital mortality, 0), there were 93.5% and 93.9% correctly classified cases in the derivation and validation cohort, respectively, whereas for the category of deceased patients (in-hospital mortality, 1), 71.0% and 72.5%. The overall proportion of correctly classified cases is 82.3% and 83.2%. The crude mortality rates and mortality-

related OR were similar in the derivation and validation cohort. The confidence intervals of OR are enlarged and statistically significant difference decreased for ORs between very-high-risk and high-risk groups as well as between low-risk and very-low-risk groups in both cohorts. The area under the ROC (95% confidence interval) was 0.823 (0.792; 0.854) for the derivation cohort (sensitivity, 0.710; specificity, 0.935) and 0.832 (0.789; 0.876) for the validation cohort (sensitivity, 0.725; specificity, 0.939). The Brier's score was estimated as 0.091 and 0.084 in the derivation and validation cohort.			
Describe any participants who were excluded from the analysis: It doesn't mention the characteristics of the excluded subjects			
Describe missing data on predictors and outcomes as well as methods used for missing data: No imputation technique of the missing values and no surrogate variables were used while tree growing. Thus, we exclude 223 (5.4%) of 4153 records from the analysis due to missing information regarding any of the predictive variables used in the classification tree.			
		Dev	Val
4.1 Were there a reasonable number of participants with the outcome?		PY	PY
4.2 Were continuous and categorical predictors handled appropriately?		Y	Y
4.3 Were all enrolled participants included in the analysis?		N	N
4.4 Were participants with missing data handled appropriately?		PY	PY
4.5 Was selection of predictors based on univariable analysis avoided?		Y	Y
4.6 Were complexities in the data (e.g. censoring, competing risks, sampling of controls) accounted for appropriately?		PY	PY
4.7 Were relevant model performance measures evaluated appropriately?		PY	PY
4.8 Were model overfitting and optimism in model performance accounted for?		NI	NI
4.9 Do predictors and their assigned weights in the final model correspond to the results from the reported multivariable analysis?		Y	Y
Risk of bias introduced by the analysis	RISK: (low/ high/ unclear)	Low	Low
Rationale of bias rating: An adequate analysis of the predictive model was made, the outcome and performance coincides with those sought in the systematic review. There are no data on model calibration.			

Step 4: Overall assessment

Use the following tables to reach overall judgments about risk of bias and concerns for applicability of the prediction model evaluation (development and/or validation) across all assessed domains.	
Reaching an overall judgment about risk of bias of the prediction model evaluation	
Low risk of bias	If all domains were rated low risk of bias. If a prediction model was developed without any external validation, and it was rated as <u>low risk of bias for all domains</u> , consider downgrading to high risk of bias . Such a model can only be considered as low risk of bias, if the development was based on a very large data set and included some form of internal validation.
High risk of bias	If at least one domain is judged to be at high risk of bias .
Unclear risk of bias	If an unclear risk of bias was noted in at least one domain and it was low risk for all other domains.
Complete for each evaluation of a distinct model.	
Reaching an overall judgment about applicability of the prediction model evaluation	

Low concerns for applicability	If low concerns for applicability for all domains, the prediction model evaluation is judged to have low concerns for applicability .
High concerns for applicability	If high concerns for applicability for at least one domain, the prediction model evaluation is judged to have high concerns for applicability .
Unclear concerns for applicability	If unclear concerns (but no “high concern”) for applicability for at least one domain, the prediction model evaluation is judged to have unclear concerns for applicability overall.

Overall judgment about risk of bias and applicability of the prediction model evaluation		
Overall judgment of risk of bias	RISK: (low/ high/ unclear)	
<i>Summary of sources of potential bias: Low</i>		
Overall judgment of applicability	CONCERN: (low/ high/ unclear)	
<i>Summary of applicability concerns: Low</i>		

38. Yagyu T, et al:

Step 2: Classify the type of prediction model evaluation

Classify the evaluation based on its aim			
Type of prediction study	PROBAST boxes to complete	Tick as appropriate	Definitions for type of prediction model study
Development only	Dev		Prediction model development without external validation. These studies may include internal validation methods such as bootstrapping and cross-validation techniques
Development and validation	Dev and Val	X	Prediction model development combined with external validation in other participants in the same article
Validation only	Val		External validation of existing (previously developed) model in other participants
<i>This table should be completed once for each publication being assessed and for each relevant outcome in your review.</i>			
Publication reference	Yagyu T, Kumada M, Nakagawa T. Novel risk stratification with time course assessment of in-hospital mortality in patients with acute heart failure. PLoS ONE 12(11): e0187410.		
Models of	Novel risk stratification with time course assessment of in-hospital mortality in patients with acute heart failure.		

interest	
Outcome of interest	Risk stratification for elderly patients with acute decompensated heart failure (ADHF)

Step 3: Assess risk of bias and applicability

DOMAIN 1: Participants			
A. Risk of Bias			
<i>Describe the sources of data and criteria for participant selection: This retrospective, observational, and single-center study was conducted between January 2009 and December 2013, enrolling a total of 1,062 consecutive AHF patients admitted to our hospital. A clinical diagnosis of AHF was made based on clinical presentation, echocardiography assessment, and laboratory testing of natriuretic peptides, and was additionally confirmed by clinical records based on Framingham criteria for the diagnosis of heart failure</i>			
	De v	Val	
1.1 Were appropriate data sources used, e.g. cohort, RCT or nested case-control study data?	Y	Y	
1.2 Were all inclusions and exclusions of participants appropriate?	Y	Y	
Risk of bias introduced by selection of participants	RISK: (low/ high/ unclear)	Low	Low
<i>Rationale of bias rating: Applied correct criteria for inclusion of participants</i>			
B. Applicability			
<i>Describe included participants, setting and dates: Of 1,062 AHF patients, 27 were excluded according to the above-mentioned criteria, with a final total of 1,035 patients enrolled. Of these, 597 patients who were admitted during or before December 2011 constituted the derivation cohort, and 438 patients who were admitted during or after January 2012 constituted the validation cohort.</i>			
Concern that the included participants and setting do not match the review question	CONCERN: (low/ high/ unclear)	Low	Low
<i>Rationale of applicability rating: The characteristics of the study population are consistent with the criteria of the research question.</i>			

DOMAIN 2: Predictors			
A. Risk of Bias			
<i>List and describe predictors included in the final model, e.g. definition and timing of assessment: BUN _ 60 mg/dL, Catecholamine Administration, Platelets < 15.0×104/μL, Total Bilirubin _ 1.2 mg/dL, CRP _ 3.0 mg/dL. For risk assessments at admission, we utilized Get With the Guidelines-Heart Failure (GWTG-HF) risk scores. We examined significant predictors of in-hospital mortality from 11 variables obtained during hospitalization and developed a risk stratification model using multiple logistic regression analysis.</i>			
	De v	Val	
2.1 Were predictors defined and assessed in a similar way for all participants?	Y	Y	
2.2 Were predictor assessments made without knowledge of outcome data?	Y	Y	
2.3 Are all predictors available at the time the model is intended to be used?	Y	Y	
Risk of bias introduced by predictors or their assessment	RISK: (low/ high/ unclear)	Low	Low

<i>Rationale of bias rating: It has a complete description of the list of selected predictor variables, as well as the way in which they were studied in patients.</i>			
B. Applicability			
Concern that the definition, assessment or timing of predictors in the model do not match the review question	CONCERN: (low/ high/ unclear)	Low	Low
<i>Rationale of applicability rating: The evaluation time of the predictive variables and the applicability of these are consistent with the research question of the recent systematic review.</i>			
DOMAIN 3: Outcome			
A. Risk of Bias			
<i>Describe the outcome, how it was defined and determined, and the time interval between predictor assessment and outcome determination:</i> The primary outcome was defined as all-cause in-hospital mortality. For the assessment of the time course of disease after admission, we collected clinical data from electronic medical records, not only at admission points but also over entire periods during hospitalization. To establish a novel predictive model, we divided the overall data into derivation and validation cohorts according to their admission dates. The former consisted of patients who were admitted up until December 2011, and the latter consisted of those who were admitted in January 2012 and there after.			
		Dev	Val
3.1 Was the outcome determined appropriately?		Y	Y
3.2 Was a pre-specified or standard outcome definition used?		Y	Y
3.3 Were predictors excluded from the outcome definition?		Y	Y
3.4 Was the outcome defined and determined in a similar way for all participants?		Y	Y
3.5 Was the outcome determined without knowledge of predictor information?		Y	Y
3.6 Was the time interval between predictor assessment and outcome determination appropriate?		Y	Y
Risk of bias introduced by the outcome or its determination	RISK: (low/ high/ unclear)	Low	Low
<i>Rationale of bias rating: Adequate description of the outcomes and the manner in which they were collected.</i>			
B. Applicability			
<i>At what time point was the outcome determined: It was determined during the follow-up in the future, during the time of hospitalization.</i>			
<i>If a composite outcome was used, describe the relative frequency/distribution of each contributing outcome:</i>			
Concern that the outcome, its definition, timing or determination do not match the review question	CONCERN: (low/ high/ unclear)	Low	Low
<i>Rationale of applicability rating: The time in which the outcome is determined is consistent with the proposed research question.</i>			
DOMAIN 4: Analysis			
Risk of Bias			
<i>Describe numbers of participants, number of candidate predictors (for DEV only), outcome events and events per candidate predictor (for DEV only):</i> Of 1,062 AHF patients, 27 were excluded according to the above-mentioned criteria, with a final total of 1,035 patients enrolled. Of these, 597 patients who were admitted during or before December 2011 constituted the derivation cohort, and 438 patients who were admitted during or after January 2012 constituted the validation cohort. Events: In derivation cohort 86 (8.3%). In validation cohort 47 (7.8%). 12 candidate parameters. EPV: derivation cohort: 7.1 validation cohort: 3.91			

Describe how the model was developed (predictor selection, optimism, risk groups, model performance): For an initial assessment at admission, we utilized an existing score model: Get With The Guidelines[®]Heart Failure (GWTG[®]DHF) risk scoring model, which is a well validated, widely accepted, and user-friendly scoring system for the risk stratification of the in-hospital mortality of AHF. Here, we excluded the parameter of ^aRace^o from this assessment because the population in this study consisted of a single race. For a time course assessment after admission, we focused on the following parameters, which could dynamically represent the condition of patients' organs and were considered as possible predictors: the presence of intravenous administration of any inotropic agent such as dopamine, dobutamine, norepinephrine or epinephrine, the presence of respiratory support by positive airway pressure with or without an invasive procedure, maximal values of total bilirubin, blood urea nitrogen, creatinine, C-reactive protein level, white blood cell concentration, and minimum values of albumin, sodium, and hemoglobin levels and platelet concentration during blood testing in all in-hospital periods, except for the day of discharge or death. To derive a simple model in the derivation cohort, continuous variables obtained after admission were converted into binary values based on optimal cut-off points, which were defined by drawing receiver operating characteristic (ROC) curves and assigning round numbers close to Youden's indexes. From 12 candidate parameters, the GWTG[®]DHF risk score at admission and the above 11 binary values obtained during hospitalization, we selected predictive variables for the best model by a stepwise selection with a P value of 0.20 for a backward elimination method, and assigned proper scores on the basis of weights of beta covariates calculated by multivariate logistic regression analysis in the derivation cohort. In this analysis, missing data were imputed by their median values. For time course assessments, we defined total scores of assigned points in selected variables at each time point as a ^atime course risk score^o. Finally, we established a summed score of the initial risk score (GWTG[®]DHF risk score) and the maximal time course risk score as a ^atotal in-hospital risk score^o. In order to validate the model performance of the total in-hospital risk score, we drew ROC curves and compared values of c-statistics in this score model and in the GWTG[®]DHF risk score model by methodology previously described by DeLong et al. Calibration of the novel model was also assessed by a Hosmer[®]DLemeshow test and visual plotting. Performances in particular subgroups were also examined in an overall cohort. Continuous variables were expressed as means $\pm$ standard deviations and compared by a Student's t test for normal distribution, or expressed as medians with interquartile ranges and compared by a Mann[®]Whitney U test for skewed distributions. Categorical variables were reported as numbers and percentages, and compared by a chi-square test or Fisher's exact test as appropriate. For all tests, a p value < 0.05 was considered statistically significant. Data were analyzed with R software packages Version 3.0.2 and JMP software Version 8.0.2(SAS Institute, Inc., Cary, NC, USA).

Describe whether and how the model was validated, either internally (e.g. bootstrapping, cross validation, random split sample) or externally (e.g. temporal validation, geographical validation, different setting, different type of participants): The article mentions the use of a derivation cohort and a validation cohort for internal and external validation respectively.

Describe the performance measures of the model, e.g. (re)calibration, discrimination, (re)classification, net benefit: An addition of time course risk score to the GWTG[®]DHF risk score significantly improved the c-statistics of the risk model for derivation and validation cohorts from 0.776 (95% confidence interval [CI]: 0.701 $\pm$ 0.852) to 0.888 (95% CI: 0.837 $\pm$ 0.938, $p = 0.002$), and from 0.806 (95% CI: 0.737 $\pm$ 0.875) to 0.902 (95% CI: 0.858 $\pm$ 0.945, $p < 0.001$), respectively (Fig 2A). The cut-off point for in-hospital mortality, which was defined on the basis of Youden's index, in the derivation cohort was 64, with a sensitivity of 0.766 and a specificity of 0.871. For the validation cohort, the sensitivity of the same score point was 0.795 and its specificity was 0.855. For the calibration of this model, plots of predicted and observed in-hospital mortalities in each decile are shown in Fig 2B. Hosmer[®]DLemeshow chi-square statistics for the two cohorts were 6.049 ($p = 0.642$) and 5.993 ($p = 0.648$), respectively. The c-statistics of subgroups for all study patients are shown in S3 Table. A significant improvement in discriminative ability, compared with the GWTG[®]DHF risk score model, was observed for most of the subgroups. Actual patients' distributions of GWTG[®]DHF, time course and total in-hospital risk scores for the overall cohort are shown in S1A $\pm$ S1C Fig. We divided the overall cohort into three groups according to GWTG[®]DHF (low: ≤ 31 , intermediate: 32 to 46, high: ≥ 47) and time course risk scores (low: ≤ 5 , intermediate: 10 to 20, high: ≥ 25). The in-hospital mortality rate for each risk status at admission and during hospitalization is presented in Fig 3A. In each risk group at admission, the in-hospital mortality rate in the high time course risk score group was higher than in that in the low and intermediate time course risk score groups (initial low risk group: 11.8% (2/17) versus 0.46% (1/218) [$p = 0.014$], initial intermediate risk group: 21.7% (23/106) versus

2.9% (16/557) [$p < 0.001$], initial high risk group: 57.6% (38/66) versus 8.5% (6/71) [$p < 0.001$]).			
<p><i>Describe any participants who were excluded from the analysis:</i></p> <p>The following AHF patients were excluded from this study: patients transferred to other hospitals for additive and urgent treatment, such as surgical intervention or heart transplantation; patients with infective endocarditis; patients with cardiopulmonary arrest on arrival; patients requiring non-pharmacological treatment for circulation support, such as intra-aortic balloon pumping and extracorporeal membrane oxygenation; and patients who were discharged or who died on the first day of admission. 27 were excluded according to the above-mentioned criterion.</p> <p><i>Describe missing data on predictors and outcomes as well as methods used for missing data: Of all variables used to calculate a new risk model, missing values were as follows: 15 values of minimum albumin levels and five values of maximum total bilirubin levels. Baseline data used to calculate the GWTG-DHF risk score did not have missing data. In this analysis, missing data were imputed by their median values.</i></p>			
	Dev	Val	
4.1 Were there a reasonable number of participants with the outcome?	N	N	
4.2 Were continuous and categorical predictors handled appropriately?	Y	Y	
4.3 Were all enrolled participants included in the analysis?	N	N	
4.4 Were participants with missing data handled appropriately?	Y	Y	
4.5 Was selection of predictors based on univariable analysis avoided?	Y	Y	
4.6 Were complexities in the data (e.g. censoring, competing risks, sampling of controls) accounted for appropriately?	Y	Y	
4.7 Were relevant model performance measures evaluated appropriately?	Y	Y	
4.8 Were model overfitting and optimism in model performance accounted for?	NI	NI	
4.9 Do predictors and their assigned weights in the final model correspond to the results from the reported multivariable analysis?	Y	Y	
Risk of bias introduced by the analysis	RISK: (low/ high/ unclear)	Low	Low
<p><i>Rationale of bias rating: An adequate analysis of the predictive model was made, the outcome and performance coincides with those sought in the systematic review.</i></p>			

Step 4: Overall assessment

Use the following tables to reach overall judgments about risk of bias and concerns for applicability of the prediction model evaluation (development and/or validation) across all assessed domains.

Reaching an overall judgment about risk of bias of the prediction model evaluation	
Low risk of bias	<p>If all domains were rated low risk of bias.</p> <p>If a prediction model was developed without any external validation, and it was rated as <u>low risk of bias for all domains</u>, consider downgrading to high risk of bias. Such a model can only be considered as low risk of bias, if the development was based on a very large data set and included some form of internal validation.</p>
High risk of bias	If at least one domain is judged to be at high risk of bias .
Unclear risk of bias	If an unclear risk of bias was noted in at least one domain and it was low risk for all other domains.

Complete for each evaluation of a distinct model.

Reaching an overall judgment about applicability of the prediction model evaluation

Low concerns for applicability	If low concerns for applicability for all domains, the prediction model evaluation is judged to have low concerns for applicability .
High concerns for applicability	If high concerns for applicability for at least one domain, the prediction model evaluation is judged to have high concerns for applicability .
Unclear concerns for applicability	If unclear concerns (but no “high concern”) for applicability for at least one domain, the prediction model evaluation is judged to have unclear concerns for applicability overall.

Overall judgment about risk of bias and applicability of the prediction model evaluation		
Overall judgment of risk of bias	RISK: (low/ high/ unclear)	
<i>Summary of sources of potential bias: Low</i>		
Overall judgment of applicability	CONCERN: (low/ high/ unclear)	
<i>Summary of applicability concerns: Low</i>		

39. Zaninovic- Jurjevic T, et al:

Step 2: Classify the type of prediction model evaluation

Classify the evaluation based on its aim			
Type of prediction study	PROBAST boxes to complete	Tick as appropriate	Definitions for type of prediction model study
Development only	Dev		Prediction model development without external validation. These studies may include internal validation methods such as bootstrapping and cross-validation techniques
Development and validation	Dev and Val	X	Prediction model development combined with external validation in other participants in the same article
Validation only	Val		External validation of existing (previously developed) model in other participants
<i>This table should be completed once for each publication being assessed and for each relevant outcome in your review.</i>			
Publication reference	Zaninović-Jurjević T, Dvornik S, Kovačić S, Kaštelan ZM, Brumini G, Matana A, et al. A simple prognostic model for assessing in-hospital mortality risk in patients with acutely decompensated heart failure, Acta Clinica Belgica, 2018. DOI: 10.1080/17843286.2018.1483562.		
Models of	A simple prognostic model for assessing in-hospital mortality risk in patients with acutely decompensated heart failure		

interest	
Outcome of interest	Combine the already known indicators with proven predictive value to develop a simple formula for assessing the risk of in-hospital mortality of patients with ADHF, as the most common form of AHF.

Step 3: Assess risk of bias and applicability

DOMAIN 1: Participants			
A. Risk of Bias			
<i>Describe the sources of data and criteria for participant selection: Participants data were collected in derivation group and validation group. Derivation patients with ADHF were considered in the period of 3 years, according to ESC guidelines. Validation patients were patients with ADHF hospitalized during 1 year of follow up. Both groups had the same inclusion and exclusion criteria.</i>			
		De v	Val
1.1 Were appropriate data sources used, e.g. cohort, RCT or nested case-control study data?		Y	Y
1.2 Were all inclusions and exclusions of participants appropriate?		Y	Y
Risk of bias introduced by selection of participants	RISK: (low/ high/ unclear)	low	Low
<i>Rationale of bias rating:</i>			
Inclusion and exclusion criteria were set for both groups, so it is secure extration information.			
B. Applicability			
<i>Describe included participants, setting and dates: The study population consisted of consecutive patients with AHF presenting to the investigator's ED.</i>			
Concern that the included participants and setting do not match the review question	CONCERN: (low/ high/ unclear)	Low	Low
<i>Rationale of applicability rating: Despite we do not know the dates, we considered that the real matter is they are collected.</i>			

DOMAIN 2: Predictors			
A. Risk of Bias			
<i>List and describe predictors included in the final model, e.g. definition and timing of assessment: It was developed a final formula that includes age, heart rate, systolic blood pressure, urea, sodium and uric acid, and this formula has a higher discriminative capacity in comparison with another group of variables, making this formula easier in applicability.</i>			
		De v	Val
2.1 Were predictors defined and assessed in a similar way for all participants?		Y	Y
2.2 Were predictor assessments made without knowledge of outcome data?		Y	Y
2.3 Are all predictors available at the time the model is intended to be used?		Y	Y
Risk of bias introduced by predictors or their assessment	RISK: (low/ high/ unclear)	Low	Low
<i>Rationale of bias rating: The article has a good selection of variables combining variables known as risk factors in ADHF.</i>			
B. Applicability			

Concern that the definition, assessment or timing of predictors in the model do not match the review question	CONCERN: (low/ high/ unclear)	Low	Low
<i>Rationale of applicability rating: They use a simple group of variables, easy to get for develop one usefull and very applicable model.</i>			
DOMAIN 3: Outcome			
A. Risk of Bias			
Describe the outcome, how it was defined and determined, and the time interval between predictor assessment and outcome determination: combine the already known indicators with proven predictive value to develop a simple formula for assessing the risk of in-hospital mortality of patients with ADHF, as the most common form of AHF.			
		Dev	Val
3.1 Was the outcome determined appropriately?		Y	Y
3.2 Was a pre-specified or standard outcome definition used?		Y	Y
3.3 Were predictors excluded from the outcome definition?		Y	Y
3.4 Was the outcome defined and determined in a similar way for all participants?		Y	Y
3.5 Was the outcome determined without knowledge of predictor information?		Y	Y
3.6 Was the time interval between predictor assessment and outcome determination appropriate?		Y	Y
Risk of bias introduced by the outcome or its determination	RISK: (low/ high/ unclear)	Low	Low
<i>Rationale of bias rating: It was very well defined outcome, with proven predictive value of one formula.</i>			
B. Applicability			
<i>At what time point was the outcome determined: at one time in the future</i>			
<i>If a composite outcome was used, describe the relative frequency/distribution of each contributing outcome:</i>			
Concern that the outcome, its definition, timing or determination do not match the review question	CONCERN: (low/ high/ unclear)	Low	Low
<i>Rationale of applicability rating: It was demonstrated that this formula is better than others in-hospital mortality risk because its simpler.</i>			
DOMAIN 4: Analysis			
Risk of Bias			
Describe numbers of participants, number of candidate predictors (for DEV only), outcome events and events per candidate predictor (for DEV only): 654 patients. This article guided its predictors from another study. Candidate predictors are not reported so the EPV cannot be estimated.			
Describe how the model was developed (predictor selection, optimism, risk groups, model performance): these two formulas (AHRSBP and urea-Na) were combined into one, with the addition of UA, whose value, expressed in $\mu\text{mol/L}$, was divided by 100 (the same reason as for Na). Thus, a final formula (AHRSBP + urea-Na + UA) was formed: $([\text{age}/10]^2 \times \text{HR}/\text{SBP}) + (\text{urea-Na}/10) + \text{UA}/100$ (where urea and Na are in mmol/L , UA in $\mu\text{mol/L}$). Final formula had a significantly greater AUC (0.741; 95% CI 0.701–0.776) in relation to its components. At the cut-off point of 53 for formula result, sensitivity of final formula was 54%, while specificity was 83%			

Describe whether and how the model was validated, either internally (e.g. bootstrapping, cross validation, random split sample) or externally (e.g. temporal validation, geographical validation, different setting, different type of participants): They validated the formula in the validation group by testing it. Patients with a higher result of the final formula had a higher mortality rate			
Describe the performance measures of the model, e.g. (re)calibration, discrimination, (re)classification, net benefit: The calibration is not available. The discrimination is given by AUC numbers in the table 3. They described the AUC with 95% CI of 0.741 for the final formula, comparing it with AHRSBP, Urea-Na/10 and Uric Acid with values of 0.717, 0.694 and 0.947 respectively. The reclassification or net benefit were not shown.			
Describe any participants who were excluded from the analysis: other causes of AHF (acute coronary syndrome, cardiogenic shock, pulmonary oedema, AHF with arterial hypertension and isolated right heart failure, death in the Emergency Unit (i.e. prior to the arrival of patients to the Department of Cardiology), heavy (dominant) chronic obstructive pulmonary disease (COPD), severe renal dysfunction (patients on dialysis treatment and those with uraemia) and history of symptomatic hyperuricemia.			
Describe missing data on predictors and outcomes as well as methods used for missing data: Do not present any missing data			
		Dev	Val
4.1 Were there a reasonable number of participants with the outcome?		Y	Y
4.2 Were continuous and categorical predictors handled appropriately?		Y	Y
4.3 Were all enrolled participants included in the analysis?		Y	Y
4.4 Were participants with missing data handled appropriately?		Y	Ni
4.5 Was selection of predictors based on univariable analysis avoided?		Y	Ni
4.6 Were complexities in the data (e.g. censoring, competing risks, sampling of controls) accounted for appropriately?		Y	Ni
4.7 Were relevant model performance measures evaluated appropriately?		Y	Y
4.8 Were model overfitting and optimism in model performance accounted for?		Y	Y
4.9 Do predictors and their assigned weights in the final model correspond to the results from the reported multivariable analysis?		Y	Y
Risk of bias introduced by the analysis		RISK: (low/ high/ unclear)	Low Unclear
Rationale of bias rating: Describe correctly the way they chose the formula and its performance.			

Step 4: Overall assessment

Use the following tables to reach overall judgments about risk of bias and concerns for applicability of the prediction model evaluation (development and/or validation) across all assessed domains.

Reaching an overall judgment about risk of bias of the prediction model evaluation

Low risk of bias	If all domains were rated low risk of bias. If a prediction model was developed without any external validation, and it was rated as <u>low risk of bias for all domains</u> , consider downgrading to high risk of bias . Such a model can only be considered as low risk of bias, if the development was based on a very large data set and included some form of internal validation.
High risk of bias	If at least one domain is judged to be at high risk of bias .
Unclear risk of bias	If an unclear risk of bias was noted in at least one domain and it was low risk for all other domains.

Complete for each evaluation of a distinct model.

Reaching an overall judgment about applicability of the prediction model evaluation

Low concerns for applicability	If low concerns for applicability for all domains, the prediction model evaluation is judged to have low concerns for applicability .
High concerns for applicability	If high concerns for applicability for at least one domain, the prediction model evaluation is judged to have high concerns for applicability .
Unclear concerns for applicability	If unclear concerns (but no “high concern”) for applicability for at least one domain, the prediction model evaluation is judged to have unclear concerns for applicability overall.

Overall judgment about risk of bias and applicability of the prediction model evaluation

Overall judgment of risk of bias	RISK: (low/ high/ unclear)	Unclear
<i>Summary of sources of potential bias: Low</i>		
Overall judgment of applicability	CONCERN: (low/ high/ unclear)	Low
<i>Summary of applicability concerns: Low</i>		