American Journal of Medicine Open 10 (2023) 100044

Contents lists available at ScienceDirect

American Journal of Medicine Open

journal homepage: www.elsevier.com/locate/ajmo

Review

Risk Prediction Models for Hospital Mortality in General Medical Patients: A Systematic Review

Yousif M. Hydoub[a], Andrew P. Walker[b,c], Robert W. Kirchoff[b,d], Hossam M. Alzu'bi[e], Patricia Y. Chipi[f], Danielle J. Gerberi[g], M. Caroline Burton[d], M. Hassan Murad[h], Sagar B. Dugani[d,i,*]

[a] Division of Cardiology, Sheikh Shakhbout Medical City, Abu Dhabi, United Arab Emirates
[b] Division of Hospital Internal Medicine, Mayo Clinic, Phoenix, Ariz
[c] Department of Critical Care Medicine, Mayo Clinic, Phoenix, Ariz
[d] Division of Hospital Internal Medicine, Mayo Clinic, Rochester, Minn
[e] Department of Cardiovascular Medicine, Mayo Clinic, Rochester, Minn
[f] Division of Hospital Internal Medicine, Mayo Clinic, Jacksonville, Fla
[g] Saint Marys Staff Library, Mayo Clinic, Rochester, Minn
[h] Evidence-Based Practice Center, Mayo Clinic, Rochester, Minn
[i] Division of Health Care Delivery Research, Kern Center for the Science of Health Care Delivery, Mayo Clinic, Rochester, Minn

ARTICLE INFO

Keywords:
Alert systems
Early warning scores
Hospital medicine
Hospital mortality
Internal medicine
Prediction models

ABSTRACT

Objective: To systematically review contemporary prediction models for hospital mortality developed or validated in general medical patients.

Methods: We screened articles in five databases, from January 1, 2010, through April 7, 2022, and the bibliography of articles selected for final inclusion. We assessed the quality for risk of bias and applicability using the Prediction Model Risk of Bias Assessment Tool (PROBAST) and extracted data using the Critical Appraisal and Data Extraction for Systematic Reviews of Prediction Modelling Studies (CHARMS) checklist. Two investigators independently screened each article, assessed quality, and extracted data.

Results: From 20,424 unique articles, we identified 15 models in 8 studies across 10 countries. The studies included 280,793 general medical patients and 19,923 hospital deaths. Models included 7 early warning scores, 2 comorbidities indices, and 6 combination models. Ten models were studied in all general medical patients (general models) and 7 in general medical patients with infection (infection models). Of the 15 models, 13 were developed using logistic or Poisson regression and 2 using machine learning methods. Also, 4 of 15 models reported on handling of missing values. None of the infection models had high discrimination, whereas 4 of 10 general models had high discrimination (area under curve >0.8). Only 1 model appropriately assessed calibration. All models had high risk of bias; 4 of 10 general models and 5 of 7 infection models had low concern for applicability for general medical patients.

Conclusion: Mortality prediction models for general medical patients were sparse and differed in quality, applicability, and discrimination. These models require hospital-level validation and/or recalibration in general medical patients to guide mortality reduction interventions.

List of Abbreviations: AIC, Akaike Information Criterion; AICC, Corrected Akaike Information Criterion; AUC, Area Under Curve; BIC, Bayesian Information Criterion; CHARMS, Critical Appraisal and Data Extraction for Systematic Reviews of Prediction Modelling studies; COVID-19, Coronavirus disease from severe acute respiratory syndrome coronavirus 2; CURB-65, Confusion, Blood urea nitrogen, Respiratory rate, Blood pressure, and Age 65; EWS, Early Warning Score; HIV, Human Immunodeficiency Virus; ICD, International Classification of Diseases; ICU, Intensive Care Unit; MEDS, Mortality in Emergency Department Sepsis; MEWS, Modified Early Warning Score; NEWS, National Early Warning Score; PIRO, Predisposition, Infection/Insult, Response and Organ dysfunction; PRISMA, Preferred Reporting Items for Systematic Reviews and Meta-Analyses; PROBAST, Prediction Model Risk of Bias Assessment Tool; qSOFA, Quick Sequential Organ Failure Assessment; SIRS, Systemic Inflammatory Response Syndrome; suPAR, Soluble Urokinase Plasminogen Activator Receptor; UVA, Universal Vital Assessment score.

* Corresponding Author: Division of Hospital Internal Medicine, Division of Health Care Delivery Research, Kern Center for the Science of Health Care Delivery, Mayo Clinic, Rochester, Minn, Tel: 1.507.255.8043, Fax: 1.507.255.9189.

E-mail address: dugani.chandrasagar@mayo.edu (S.B. Dugani).

https://doi.org/10.1016/j.ajmo.2023.100044
Received 13 August 2022; Accepted 27 May 2023
Available online 5 June 2023

Introduction

The global burden of hospital mortality from preventable and nonpreventable causes is high.[1] Recent studies estimated that 3.1% of hospital deaths were preventable[2] and that 134 million adverse events annually in low- and middle-income countries resulted in 2.6 million hospital deaths.[1,3] Despite these and other studies, there is sparse information on risk prediction models for preventable hospital mortality, limiting the development of mortality reduction interventions. Furthermore, predicting nonpreventable deaths may facilitate earlier discussion of advanced care directives or transition to palliative or hospice care.

In the United States, from 2000 to 2010, the annual overall hospital mortality rate declined from 2.5/100 patients to 2.0/100 patients.[4] While reassuring, the overall rate concealed changes in mortality were attributed to different conditions, such as kidney disease (–65%), heart disease (–16%), and septicemia (+17%).[4] In this context, disease-specific models have been developed to predict risk of hospital mortality for conditions including acute myocardial infarction, stroke, heart failure, pancreatitis, and pneumonia.[5-9] In contrast to disease-specific models, there are fewer models for general medical patients, who frequently present with undifferentiated and/or multiple medical conditions and experience an unpredictable hospital course. Furthermore, general medical patients have large amounts of evolving biopsychosocial data that are challenging to integrate in prediction models for mortality. Therefore, early identification of general medical patients at risk of hospital mortality may improve decision making, guide escalation in care, and reduce the risk of preventable deaths. Despite this, there is sparse information on risk prediction models for hospital mortality for general medical patients, and specifically, no systematic appraisal of model quality and performance.

To address these knowledge gaps, we conducted a systematic review of risk prediction models for hospital mortality in general medical patients. The objective of the study was to evaluate models that predicted acute decompensation, focusing on deaths in general medical patients. We reviewed model characteristics and performance and critically appraised their quality and performance. These models may be validated, recalibrated, or improved to guide interventions to reduce hospital mortality.

Methods

The study was conducted by the Hospital Experiences to Advance Goals and Outcomes Network (HEXAGON) group at Mayo Clinic.[10-12] The systematic review is reported following the Preferred Reporting Items for Systematic Reviews and Meta-analyses (PRISMA) guidelines and the protocol was prospectively registered (PROSPERO CRD42020176054).[13,14] The search strategy for all databases is provided in Supplemental Table 1 (available online).

Data Sources and Search Methods

We searched the peer-reviewed literature for articles on models to predict hospital mortality in general medical patients. We searched Ovid/Medline, Embase, Evidence-Based Medicine Reviews, Scopus, and Web of Science from January 1, 2010, through April 7, 2022. We restricted the search period to focus on models reflecting contemporary cohorts and hospital practice. The bibliography of articles selected for final inclusion was screened by one investigator.

Study Selection

We defined hospital mortality as death occurring in the hospital following admission to a general medical ward. We included original English language articles that reported at least one prediction model for mortality in adults (age ≥ 18 years) hospitalized on general medical wards. We focused on models designed to predict acute decompensation, focusing on mortality. We excluded studies with any of the following: patients admitted from the emergency department to the intensive care unit (ICU); patients on surgical, palliative care, oncology, or cardiology services; and patients with coronavirus disease from severe acute respiratory syndrome coronavirus 2 (COVID-19). We excluded studies that focused on one diagnosis/condition (eg, pneumonia). To avoid overfitting of models, we excluded studies with fewer than 500 mortality events. We also excluded studies that focused on specific time horizons for hospital mortality (eg, 7-day mortality), as predicting mortality at any point during hospitalization may guide overall hospital care and mortality reduction interventions. After pilot screening, 2 investigators independently screened titles and/or abstracts (primary screening) and full-text (secondary screening) using Covidence.[15]

Data Extraction and Quality Assessment

We categorized the models as "general models" if developed and/or validated in all general medical patients and as "infection models" if developed and/or validated in patients admitted to medical wards with suspected infection. Data from articles selected for final inclusion were extracted at the model level using the Critical Appraisal and Data Extraction for Systematic Reviews of Prediction Modelling Studies (CHARMS) checklist.[16] Quality was assessed using Prediction Model Risk of Bias Assessment Tool (PROBAST).[17] Using PROBAST, we assessed risk of bias using 4 domains (Participants, Predictors, Outcome, and Analysis), and applicability using 3 domains (Participants, Predictors, and Outcome) (Supplemental file 2).

Conflict Resolution

Screening, data extraction, and quality assessment were conducted independently by two investigators, with conflicts resolved by discussion and consensus between the investigators or, if needed, by a third investigator.

Ethics Approval

The study used publicly available, deidentified data and forms and therefore did not require Institutional Review Board approval.

Results

Of 20,424 unique articles, 8 studies on 15 prediction models met selection criteria (Figure 1). The studies, published since 2017, were based in 10 countries across 4 world regions (Table 1 and Supplemental Table 2 [available online]). The 8 studies on 280,793 patients included 128,017 women (45.6%) and 19,923 hospital deaths (7.1%) (Table 1). Of the 15 models, 10 models were based on general medical patients, 7 models on general medical patients with infection, and 2 models were studied in both groups. Out of the 15 models, 7 were novel or modified,[18-24] while the others[19,20,24,25] were preexisting models externally validated in general medical patients. Major findings of the systematic review are summarized in Figure 2. Heterogeneity in variables and models precluded a metaanalysis.

Study Design, Cohorts, and Outcomes

The studies were based on cohorts of 5000 to more than 100,000 participants. All cohorts had a mean or median age above 60 years, except in Moore 2017,[20] which had a median age of 36 years and additionally provided subgroup analysis for general medical patients with infection. Two studies exclusively enrolled patients with infection (Chen, 2017[24] and Fabbian, 2018[25]), while the others enrolled all patients admitted to the general medical wards. One study (Soffer 2020[23]) focused on mortality in the general medical ward. The other studies focused on hospital mortality (in and outside the general medical ward) after admission to a general medical ward.

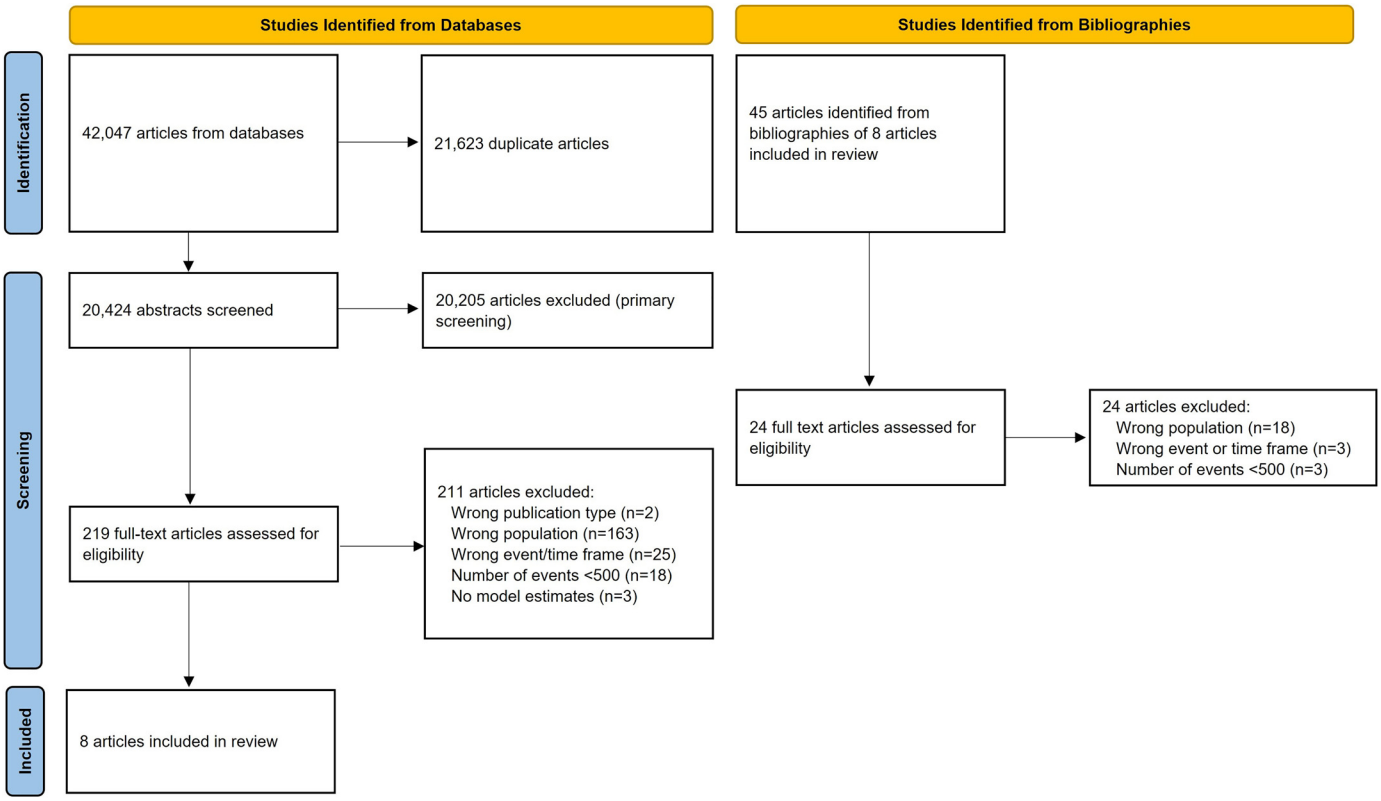

Figure 1. Preferred Reporting Items for Systematic Reviews and Meta-analyses (PRISMA) 2020 flowchart of articles included for data extraction. The search strategy is provided in Table 1 in the Supplement (available online).

Characteristics	General Models	Infection Models
Data source	• 249 499 patients and 16 754 hospital deaths • Africa (3 models), Asia (3 models), and Europe (4 models) • Study period (2000–2015) • Novel (4 models), Modified (2 models), and Preexisting (4 models)	• 34 447 patients and 3 889 hospital deaths • Africa (1 model), Asia (5 models), and Europe (1 model) • Study period (2010–2015) • Novel (2 models) and Preexisting (5 models)
Predictors	• Early Warning Scores (5 models), Comorbidities (2 models), and Combination (3 models) • Number of final predictors (3–62) • Predictors selected using P values (3 models), BIC (2 models), AIC and AICC (1 model), N/A (4 models) • Novel/modified models developed using logistic regression (3 models), Poisson regression (1 model), and machine learning (2 models)	• Early Warning Scores (3 models), Comorbidities (1 model), and Combination (3 models) • Number of final predictors (4–16) • Predictors selected using AIC (1 model), BIC (1 model), N/A (5 models • Novel models developed using logistic regression (2 models)
Missing values	• Missing values imputed (3 models), assigned value of 0 (2 models), excluded (3 models), no information (2 models)	• Missing values imputed (1 model), excluded (5 models), no information (1 model)
Calibration	• Calibration plot (none) • Hosmer–Lemeshow test (2 models) • No calibration (8 models)	• Calibration plot (1 model) • Hosmer–Lemeshow test (1 model) • No calibration (5 models)
Discrimination	• AUC: >0.8 (4 models) and 0.65–0.79 (6 models)	• AUC: >0.8 (none), 0.65–0.79 (6 models), and <0.65 (1 model)
Validation	• Temporal data split (3 models), Bootstrap (2 models), and Cross validation (1 model) • External validation of preexisting models (4 models)	• Temporal data split (1 model) and Cross validation (1 model) • External validation of preexisting models (5 models)

Figure 2. Summary of the main results of the systematic review on models for hospital mortality in general medical patients. Used with permission of Mayo Foundation for Medical Education and Research, all rights reserved.

Table 1
Characteristics of Included Studies.

Characteristic	Chen (2017)[24]	Fabbian (2017)[18]	Fabbian (2018)[25]	Moore (2017)[20]	Rasmussen (2018)[19]	Sakhnini (2017)[21]	Schwartz (2018)[22]	Soffer (2020)[23]
Study objective	Development and internal validation of a novel model, and external validation of existing models	Model modification	External validation	Development and internal validation of a novel model, and external validation of existing models	Model modification and external validation of existing models	Model development and internal validation	Model development and internal validation	Model development and internal validation
Location (number of centers)	Taiwan (1)	Italy (1)	Italy (1)	Gabon, Malawi, Sierra Leone, Tanzania, Uganda, Zambia (unknown)	Denmark (2)	Israel (1)	Israel (1)	Israel (1)
Study period*	2010-2012	2000-2013	2013-2016	2009-2015	2013-2015	2013-2015	2012-2015	Development (2013-2017), Validation (2018)
Inclusion criteria	Adults in the emergency department with ICD-9 codes for infection, and with ≥2 sets of blood cultures	Admissions to the medical ward	Admissions to the medical ward with an infectious disease diagnosis	Admissions to the medical ward with mortality data and >50% of recorded vital signs	Admissions to the medical ward and suPAR analysis	Admissions to the medical ward ≥24 h	Admissions to the medical ward ≥24 h	Patients (18-100 years) admitted to the medical ward
Exclusion criteria	Patients transferred from other medical institutions, with repeat hospital visits, and/or with traumatic injuries	Patients transferred to surgical departments	Patients transferred to surgical departments or intensive care units	Not applicable	Admissions for surgical intervention or acute pediatric, obstetric, and gastroenterological conditions; patients in whom suPAR level was not ordered, suPAR result was missing, or other reasons	Admissions classified under symptoms, signs, and ill-defined conditions (ICD-9 codes 780-799), and under observation (ICD-9 codes V71, V71.2, V29.0, V29.1, V29.2, V29.8, V29.9)	Not applicable	Not applicable
Age (years), mean ± SD, median (IQR)†	65 (49-78)	73 ± 16	65 ± 25	36 (27-49)	61 (43-76)	Survived: 68 (18-105)** Died: 78 (23-105)**	Survived: 64 (18-105)** Died: 77 (23-105)**	Survived: 73 (62-83) Died: 82 (71-89)
Number of patients	Development: 7011 Validation: 12,110	Total: 75,586	Total: 12,173	Total: 5573 With infection: 3153	Total: 17,312	Development: 7268 Validation: 7843	Development: 10,788 Validation: 6867	Total: 118,262§§
Women, number (%)	3216 (45.9)	40,329 (53.4)	8053 (66.2)	2829 (50.8)	9194 (53.1)	Survived: 2984 (46.7)†† Died: 453 (51.4)††	Survived: 4404 (44.4) Died: 450 (51.5)	Died: 3042 (48.2)
Major admitting diagnoses (%)	Unspecified infections (71.2), respiratory infections (57.9), genitourinary infections (38.2)	% not provided	Pulmonary infection (34.3), nonspecified infection (33.8), urinary tract infection (17.5)	Not provided	Not provided	Survived: nonspecified diagnosis (20.5), heart failure (12.1), cerebrovascular disease (11.4). Died: pneumonia (23.4), sepsis and septicemia (14.5), malignant neoplasms (14.1)	Survived: nonspecified diagnosis (52.2), heart failure (6.5), pneumonia (5.7). Died: nonspecified diagnosis (22.8), pneumonia (19.3), sepsis and septicemia (13.2)	Survived: nonspecified chest pain (5.8), pneumonia (5.0), CHF exacerbation (2.7). Died: septic shock (13.7), pneumonia (13.3), respiratory failure (3.9)
Number of deaths§	Development: 479 Validation: 1145	6007	1545	Total: 966 With infection: 720	587	Development: 882 Validation: 582	Development: 874 Validation: 515	6311
Mortality prediction model	CHARM, CURB-65, MEDS, PIRO, SIRS	Modified Elixhauser Index	Modified Elixhauser Index	MEWS, qSOFA, UVA	NEWS, modified NEWS	Not applicable	Not applicable	Not applicable

* All studies used single retrospective cohorts. Moore (2017)[20] used a pooled cohort.
† Rounded to integers.
§ All studies included all hospital mortalities, except Soffer (2020),[23] which included only ward mortalities.
** Mean (range).
†† From development cohort.
§§ One dataset was used with gradient boosting.
Abbreviations listed in Supplemental Table 13 (available online).

Prediction Models for All General Medical Patients

Ten of 15 models were based on general medical patients hospitalized with various medical conditions (Table 2). The Universal Vital Assessment (UVA) score[20] and 3 other models[21-23] were developed and internally validated in general medical patients.[20-23] Two models were modified versions of the Elixhauser index[18] and National Early Warning Score (NEWS).[19] The Elixhauser Index,[18] Modified Early Warning Score (MEWS),[20] NEWS,[19] and Quick Sequential Organ Failure Assessment (qSOFA) score[20] were preexisting models externally validated in general medical patients. None of the studies both developed and externally validated model(s) for mortality. Studies that developed or modified models examined different predictors for hospital mortality, with the most common categories being vital signs (5 models) and patient comorbidities (4 models) (Table 2 and Supplemental Table 3 [available online]). Based on the type of predictors, the models were classified as Early Warning Score (EWS), comorbidity, and combination models (Supplemental Table 4, available online). EWS models were generally based on vital signs, with few models also integrating easily obtainable biomarkers. Comorbidity models were typically weighted scores of comorbidities from International Classification of Diseases (ICD) codes. Combination models contained variable categories of predictors including vital signs, comorbidities, and biomarkers. Most predictors were captured on hospital presentation or admission to the medical ward, with a minority from hospital progress notes and discharge summaries.[18,21-23] Predictors in all novel and modified models were selected using multivariable analysis. The most common methods to select predictors were probability values (P values) (n = 3 models)[18,19,21]; UVA score used Bayesian information criterion (BIC)[20]; and Schwartz (2018)[22] used Akaike information criterion (AIC), corrected Akaike information criterion (AICC), and BIC (Table 2). Most general models used data in their native form, except the modified Elixhauser Index, which used age as a categorical variable (Supplemental Table 5, available online).

Most models were developed using logistic regression, and model performance was reported using discrimination. The modified Elixhauser Index,[18] UVA score,[20] and model by Sakhnini (2017)[21] were developed using logistic regression, while the model by Schwartz (2018)[22] used the least absolute shrinkage and selection operator method. The modified NEWS[19] was developed using Poisson regression, while the machine learning model by Soffer (2020)[23] employed tree-based classifiers with gradient boosting. Calibration was assessed using the Hosmer-Lemeshow test in the modified Elixhauser Index[18] and in the model by Sakhnini (2017),[21] whereas other models did not report calibration. In terms of model discrimination, NEWS,[19] modified NEWS,[19] and the models by Sakhnini (2017),[21] Schwartz (2018)[22], and Soffer (2020)[23] had high discrimination (area under curve [AUC] > 0.8), while the other models showed moderate discrimination (AUC 0.65-0.79) (Table 2). Most models[19-24] reported other measures including sensitivity and specificity (Supplemental Table 6, available online). Only the UVA score reported a subgroup analysis[20] (Supplemental Table 7, available online).

The studies differed in their handling of missing values (Table 2 and Supplemental Table 8 [available online]). After excluding participants with missing data regarding human immunodeficiency virus (HIV; n = 2171) and mortality outcomes (n = 8), MEWS, qSOFA, and UVA score in Moore (2017)[20] imputed missing values. The model by Soffer (2020)[23] used gradient boosting algorithms for missing values. Other models either excluded[19,21,22] or did not provide information[18] on handling of missing values.

Prediction Models for General Medical Patients with Infection

Table 3 outlines the characteristics of infection models. The CHARM score[24] and UVA score[20] were the only novel models developed and internally validated in general medical patients with infection. Fabbian (2018)[25] externally validated the modified Elixhauser Index, which they reported previously (Fabbian, 2017[18]). Chen 2017[24] externally validated 4 novel models: (i) Predisposition, Infection/Insult, Response and Organ dysfunction (PIRO); (ii) Confusion, blood Urea nitrogen, Respiratory rate, Blood pressure, and age 65 (CURB-65); (iii) Mortality in Emergency Department Sepsis (MEDS) score; and (iv) Systemic Inflammatory Response Syndrome (SIRS). None of the models were developed and externally validated in an independent dataset within the same study. Predictors for the novel models (CHARM and UVA) were selected using multivariable analysis. Vital signs were used as predictors in 5 models, while comorbidities were used in 3 models (Table 3 and Supplemental Table 3 [available online]). Predictors for all models were captured on admission, except for the modified Elixhauser Index,[25] which captured ICD codes from discharge summaries. The CHARM score[24] converted continuous predictors into categorical predictors (Supplemental Table 5, available online) and used AIC to select predictors. All models in Chen (2017)[24] excluded missing values. Calibration was only reported for the CHARM score and had adequate calibration based on a calibration plot and Hosmer-Lemeshow test. The CHARM score had the highest AUC of 0.77 (95% CI, 0.75-0.79) in the development cohort and 0.76 (95% CI, 0.75-0.77) in the validation cohort, while SIRS had a modest AUC of 0.58 (95% CI, 0.56-0.61). Other models had an AUC ranging from 0.68 to 0.75 (Table 3). Alternative performance measures (Supplemental Table 9, available online) and subgroup analysis (Supplemental Table 7, available online) were also reported.

Quality and Applicability Assessment

Based on PROBAST criteria, all general models had an overall high risk of bias (Supplemental Figure 1A, available online). All general models had low risk of bias for the *Participant* domain, indicating the appropriateness of the populations for our study. For the *Predictors* domain, 4 of 10 general models had a low risk of bias. Common reasons for high risk of bias were the timing of predictor measurement and availability of predictors on admission, in particular for models requiring laboratory data[19,20,22,23] or ICD codes[18,21-23] (Supplemental Table 10, available online). For the *Outcomes* domain, 5 of 10 general models had an unclear risk of bias, attributed to unclear duration from predictor measurement to hospital mortality. All general models had high risk of bias for the *Analysis* domain because none of the studies reported both model calibration and discrimination. MEWS,[20] qSOFA,[20] NEWS,[19] and the model by Soffer (2020)[23] had low concern for applicability to predict hospital mortality for all general medical patients (Supplemental Figure 1C, available online).

All infection models had an overall high risk of bias and a low risk of bias for the *Participant* domain (Supplemental Figure 1B and Supplemental Table 10, available online). For the *Predictors* domain, 5 of 7 infection models had low risk of bias. For the *Outcomes* domain, all infection models, except for the UVA score,[20] had low risk of bias. All infection models had a high risk of bias for the *Analysis* domain. Only the CHARM score[24] provided sufficient information on model calibration and discrimination and appropriately used univariate models to select predictors (Table 3). CHARM,[24] PIRO,[24] CURB-65,[24] MEDS,[24] and SIRS[24] scores had low concern for applicability as hospital mortality predictors for patients with infection in the medical ward (Supplemental Figure 1D, available online).

Discussion

In this systematic review of 8 studies on 280,793 patients from 10 countries, we identified 15 risk prediction models for hospital mortality in general medical patients. Of these models, 5 were novel, 2 were adapted, and 8 were external validation of preexisting models. For general models, NEWS and the model by Soffer (2020) had an optimal balance of discrimination, bias, and applicability. For infection models, the CHARM and PIRO scores had an optimal balance. Overall, there was sparse data on risk prediction and the available models had a high risk

Table 2
Prediction Models for All General Medical Patients (General Models)

Characteristic	Elixhauser Index[18]	Modified Elixhauser Index[18]	MEWS[20]	NEWS[19]	Modified NEWS[19]	qSOFA[20]	UVA[20]	Model by Sakhnini (2017)[21]	Model by Schwartz (2017)[22]	Model by Soffer (2020)[23]
Purpose of analysis	External validation/ comparison	Model modification	External validation	External validation/ comparison	Model modification	External validation	Model development/ internal validation	Model development/ Internal validation	Model development/ Internal validation	Model development/ Internal validation
Timing of predictor assessment	At diagnosis of comorbidity (ICD9 codes)	At diagnosis of comorbidity (ICD9 codes)	On admission	On admission	On admission	On admission	On admission	On admission	On admission	On admission
Number of candidate predictors	Not applicable	33	Not applicable	Not applicable	10	Not applicable	13	28	32	Not provided
Predictors in final model (no.)	Congestive heart failure, cardiac arrhythmias, valvular disease, pulmonary circulation disorders, peripheral vascular disorders, hypertension, paralysis and other neurological disorders, chronic pulmonary disease, diabetes mellitus, hypothyroidism, renal failure, liver disease, peptic ulcer disease excluding bleeding, HIV, lymphoma and cancer, rheumatoid arthritis/collagen vascular diseases, coagulopathy, obesity, weight loss, fluid and electrolyte disorders, anemia, alcohol and drug abuse, psychoses, depression (30)	Age, renal failure, male gender, other neurological disorders, lymphoma, solid tumor without metastasis, ischemic heart disease, congestive heart failure, coagulopathy, fluid and electrolyte disorders, liver disease, weight loss, metastatic cancer (13)	Temperature, HR, RR, SBP, level of consciousness (5)	RR, Temp, SBP, HR, AVPU, SpO_2, supplemental oxygen (7)	NEWS predictors, age, sex, suPAR level (10)	RR, SBP, level of consciousness (3)	Temperature, HR, RR, SBP, SpO_2, level of consciousness (GCS), HIV serostatus (7)	Age, BMI, mean arterial pressure on admission, previous admission within 3 prior months, heart failure, active malignancy, chronic use of statins, chronic use of antiplatelet agents, main admission diagnosis and secondary conditions (heart failure, urinary tract infection, pneumonia, sepsis and septicemia, renal failure, cancer, and acute coronary syndrome) (16)*	Age, BMI, admission within 3 prior months, statin intake, 4 laboratory variables (serum creatinine, hemoglobin, RDW, and hypoalbuminemia), and 2 background diseases (heart failure and malignancy) (10)	Chief complaint, age, home medications, number of home medications, comorbidities, number of comorbidities, RR, SpO_2, fever, SBP, DBP, pulse, albumin, BUN, CRP, LDH, neutrophil count, eosinophil count, WBC count, calcium, AST, PCO_2, lactate, protein, lymphocyte count, serum creatinine, phosphorus level, ALK-P, troponin-I, hemoglobin, potassium, glucose, sodium, PT, INR, GGT, platelet count, ED diagnosis, ED administered medications, arrival mode, ED wing, emergency severity index, pain score and textual words (catheter, shock, fluid, bedridden, oxygen, feeding tube, culture, the family, low, gas, deterioration, nursing, dementia, consciousness, ambulance, intubated, brought, condition, breathing)(62)†

(*continued on next page*)

Table 2 (*continued*)

Characteristic	Elixhauser Index[18]	Modified Elixhauser Index[18]	MEWS[20]	NEWS[19]	Modified NEWS[19]	qSOFA[20]	UVA[20]	Model by Sakhnini (2017)[21]	Model by Schwartz (2017)[22]	Model by Soffer (2020)[23]
Method for handling missing values	Not provided	Not provided	k-nearest neighbors imputation[§]	Missing vital signs assigned value of 0	Missing suPAR level values were excluded; missing vital signs assigned value of 0	k-nearest neighbors imputation[§]	k-nearest neighbors imputation§	Exclusion	Exclusion	Integrated into gradient boosting algorithm
Method for prediction model development	Not applicable	Logistic regression	Not applicable	Not applicable	Poisson regression	Not applicable	Logistic regression; decision trees and linear regression	Logistic regression	Penalized logistic regression (LASSO)	Multiple tree-based classifiers with gradient boosting
Method for selection of predictors during multivariable analysis	Not applicable	*P* value	Not applicable	Not applicable	*P* value	Not applicable	BIC	*P* value	AIC, AICC, BIC	Not applicable
Calibration method (result)	None	Hosmer-Lemeshow test ($P < .001$)	None	None	None	None	None	Hosmer-Lemeshow test (not provided)	None	None
AUC value (95% CI)[**]	0.66 (0.65-0.66)	0.72 (0.71-0.73)	0.70 (0.68-0.71)	0.87 (0.85-0.88)	0.92 (0.91-0.92)	0.69 (0.67-0.72)	0.77 (0.75-0.79)	Development: 0.90 Validation: 0.81	Development: 0.89 (0.88-0.90) Validation: 0.86 (0.84-0.87)	0.92 (0.92-0.93)
Type of validation	External	None	External	External	None	External	Internal	Internal	Internal	Internal
Method of validation	Different time, area and investigators		Different time, area and investigators	Different time, area and investigators		Different time, area and investigators	10-fold cross-validation	Temporal data split	Bootstrap and temporal data split	Bootstrap and temporal data split

[*] Obtained from Supplement.

[†] List of predictors obtained from different tables reported in the article.

[§] Variables with more than 50% missing values were excluded. Hospital mortality and HIV serostatus were not imputed.

[**] Rounded to 2 decimal places.

Abbreviations listed in Supplemental Table 13 (available online).

Table 3
Prediction Models for General Medical Patients with Infection (Infection Models)

Characteristic	CHARM[24]	PIRO[24]	CURB-65[24]	MEDS[24]	SIRS[24]	Modified Elixhauser Index[25]	UVA[20]
Purpose of analysis	Model development/internal validation	External validation/model comparison	External validation/model comparison	External validation/model comparison	External validation/model comparison	External validation	Model development/internal validation
Timing of predictor assessment	On admission	On admission	On admission	On admission	On admission	At diagnosis of comorbidity (ICD-9)	On admission
Number of candidate predictors	62	Not applicable	Not applicable	Not applicable	Not applicable	Not applicable	13
Predictors in final model (no.)	Absence of chills, anemia, hypothermia, malignancy, RDW (5)	Age, bands >5%, BUN, COPD, HR, lactate, liver disease, malignancy, nursing home residence, other infection, platelet count, pneumonia, respiratory failure/hypoxemia, RR, SBP, skin/soft tissue infection (16)*	Age, BUN, confusion, RR, SBP or DBP (5)*	Age, altered mental status, bands >5%, lower respiratory tract infections, nursing home resident, platelet count, respiratory disease, septic shock, terminal disease (9)*	HR, RR or $PaCO_2$, temperature, WBC count or bands >10% (4)*	Age, coagulopathy, congestive heart failure, fluid and electrolyte disorders, gender, ischemic heart disease, liver disease, lymphoma, metastatic cancer, other neurological disorders, renal failure, solid tumor without metastasis, weight loss (13)	HIV serostatus, HR, level of consciousness (GCS), RR, SBP, SpO_2, temperature (7)
Method for handling missing data	Variables with more than 5% missing values were not considered candidate predictors	Not applicable	Not applicable	Not applicable	Not applicable	Not provided	k-nearest neighbors imputation†
Method for prediction model development	Logistic regression	Not applicable	Not applicable	Not applicable	Not applicable	Not applicable	Logistic regression; decision trees and linear regression
Method for selection of predictors during multivariable analysis	AIC	Not applicable	Not applicable	Not applicable	Not applicable	Not applicable	BIC
Calibration method (result)	Calibration plot and Hosmer-Lemeshow test ($P = .42$)	None	None	None	None	None	None
AUC value§	Development: 0.77 (0.75-0.79); Validation: 0.76 (0.75-0.77)	0.74 (0.72-0.77)	0.68 (0.65-0.70)	0.67 (0.64-0.69)	0.58 (0.56-0.61)	0.72 (0.71-0.74)	0.75 (0.72-0.77)
Type of validation	Internal	External	External	External	External	External	Internal
Method of validation	Temporal data split	Different time, area, and investigators	Different time, area, and investigators	Different time, area, and investigators	Different time, area, and investigators	Different time and area	10-Fold cross-validation

* Predictors of PIRO, CURB-65, MEDS, SIRS were extracted from references in Chen (2017).[24]
† Variables with more than 50% missing values were excluded. Hospital mortality and HIV serostatus were not imputed.
§ Rounded to 2 decimal places.
Abbreviations listed in Supplemental Table 13 (available online).

of bias attributed, in part, to suboptimal calibration and handling of missing values. The novel models had higher discrimination compared to the adapted and validated models. To our knowledge, there is no comparable synthesis of studies on risk prediction models, and our findings identify the need for studies on risk prediction models for hospital mortality in general medical patients.

EWS models, frequently incorporated as track and trigger systems, are simple point-based systems ideal for predicting short-term mortality (eg, 24 or 48 hours) in ICU and general medical patients.[26,27] Compared with general medical wards, there was more information on predicting mortality in the ICU, which has informed studies in non-ICU settings.[28-31] Similar to ICU settings, early and prompt recognition of general medical patients at high risk for mortality is an important first step in preventing mortality.[32] Furthermore, general medical patients frequently have complex acute and chronic illnesses, and early prediction of nonpreventable deaths may facilitate their transition to palliative or hospice care. Simple EWS models (eg, MEWS, NEWS) have variable performance[27,33] and are well suited for patients with infection, who typically experience frequent change in vital signs reflecting potential clinical deterioration. A retrospective study showed that EWS models outperformed other models in patients with infections but were less promising in patients without infection.[33] Given the dependence on vital signs, EWS models are less suitable for predicting long-term hospital mortality (eg, 7 days).[26,34] Machine learning models[32,35] have been explored as alternatives to EWS models and have the advantage of automated calculations that reduce human error, improve integration in electronic health record (EHR) systems, and generate fewer false alarms. Mortality prediction models have helped to identify high-risk patients in non-ICU settings. A recent randomized controlled clinical trial showed that automated detection and monitoring of clinical deterioration in hospitalized adults was associated with a 16% reduction in 30-day mortality following an alert.[36]

In contrast to EWS models, comorbidity models including the Charlson Comorbidity Index[37] and the modified Elixhauser Comorbidity Index[38] may be better for predicting longer-term (ie, more than 48 h) hospital mortality.[39-41] Comorbidity models have applications beyond predicting hospital mortality and have been used to predict healthcare expenditure, tailor treatment, and control for confounders in epidemiologic studies.[42,43] However, their benefit to healthcare providers may be limited because most comorbidity models require knowledge of comorbidities on hospital admission, which may not be readily available for many patients. Furthermore, comorbidity models do not detect rapid deterioration in clinical status and may have high concern for applicability in hospitals with limited use of EHRs.

Real-world performance and usability of prediction models may be affected by many factors. The number of predictors in a model does not necessarily correlate with its predictive performance, as summarized in a previous systematic review.[44] In the current study, 6 of 10 general models and 2 of 7 infection models had 10 or more predictors. The number of predictors in a model can affect model performance in real-world settings, particularly in hospitals with limited resources.[45] Implementing models in hospitals with understaffed wards, scarce monitoring systems, and limited technology may increase the burden of manual labor with minimal benefit on mortality.[45-47] Further, some predictors may be relevant in a specific world region, improving its performance. As Moore (2017)[20] illustrated in sub-Saharan Africa, adding HIV serology to vital signs resulted in better discrimination of the UVA score over other EWS scores.

During screening, we excluded articles that focused on specific time horizons for hospital mortality (eg, 72-hour or 7-day mortality). While a time horizon of a few days may be appropriate for EWS models,[26] mortality at any point during hospitalization may be more relevant to guide overall hospital care that integrates biopsychosocial factors and patients' preferences.[33] Some suggest that 30-day postdischarge mortality, rather than hospital mortality, is a better indicator of hospital performance.[48,49] However, hospital mortality may be more relevant, in places with limited ability to provide, influence, or monitor postdischarge care.[50,51] We excluded models such as the eCART score, which did not meet study selection criteria.[52,53] However, the eCART score, based on medical and surgical patients, integrated vital signs and laboratory data to predict outcomes in the subsequent 24 hours. eCART and other automated scores that guide short-term care may have high clinical uptake.

Fourteen of 15 models were developed or validated in single-site studies. The most common method of internal validation in the included models was temporal data split ($n = 4$ models); however, cross-validation and bootstrapping may be preferred over random and temporal data split, particularly in smaller datasets.[54] None of the novel or modified models in general medical patients were externally validated, and preexisting models were only externally validated in 1 country. To ensure model stability and generalizability, Riley et al.[55] recommend externally validating prediction models using big datasets from individual participant data metaanalysis or EHRs that include participants from different regions and providing subgroup analysis to study geographic heterogeneity in model performance. Validating these models in regions with different demographics, patient case-mix, staffing volumes, and technologic capability may improve generalizability and guide care in hospitals that have limited ability for research and evaluation.[55,56] Additionally, in the current study, heterogeneity in population characteristics precluded direct comparison of models across studies. For instance, the population was younger in the Moore (2017) study (median age: 36 years; interquartile range: 27-49 years) compared with other studies. Thus, the generalizability of these models to other hospitals, countries, and world regions is unknown. However, studies based on local data have the ability to calibrate existing models to local demographics and patient case-mix, leading to informative models that can guide mortality reduction interventions.[55,57,58]

As highlighted, no single model will apply to all populations and healthcare settings. Thus, the resources needed to implement a prediction model would depend on the model. Machine learning models are of emerging popularity in many high-income countries. The popularity of these models stems from their high performance, abundance of EHR data, advanced technological expertise, and capability to evaluate and enhance them. Due to the simplicity of their implementation in resource-limited settings, early warning scores may be more feasible. Ultimately, these models may aid in timely escalation of care and/or transition to palliative or hospice care. Incorporating models into hospital practice will require an impact analysis on hospital mortality.[59]

Limitations and Strengths

The study has potential limitations. We restricted our search to English-language articles, which may influence the generalizability of our results, but not necessarily result in systemic bias.[60] During screening, we excluded articles that did not explicitly distinguish general medical patients from others (eg, surgical patients) and inadvertently may have excluded relevant articles. To mitigate this, we used a broad search strategy for primary screening with independent review by two investigators.[61] We excluded articles on COVID-19 infection because risk factors and treatment evolved during the pandemic, thereby influencing the risk of mortality. Therefore, our findings may not be generalizable to patients with COVID-19. Studies in our analysis were conducted in the prepandemic period; the COVID-19 pandemic has resulted in a global excess mortality of ~3 million deaths in 2020[62] and hospital mortality rates in some countries increased for non-COVID-19 patients.[63] Therefore, the performance of prediction models for non-COVID-19 patients may be different during the pandemic. The study has several strengths. Our review focused on mortality in contemporary cohorts and reflected current practice and advances in model prediction, machine learning, and artificial intelligence. We used CHARMS and PROBAST, which are rigorous tools for data extraction and quality assessment.[17] We included studies with >500 mortality events resulting in ~50 events

per predictor variable, well above the recommendation for prediction models.[26,64]

Conclusions

In this systematic review of 8 studies, 14 of 15 risk prediction models for hospital mortality were from single-site studies, which have high local relevance but unknown generalizability. All models had a high risk of bias and differed in model covariates, applicability, and discrimination. There is a need for rigorous models to predict mortality in general medical patients. Rather than disease-specific models, unified prediction models for general medical patients calibrated to the local determinants of hospital care, including patient case-mix, technologic availability, and workforce capability, may be better incorporated into clinical decision support tools and facilitate the delivery of safer hospital care.

Funding/Support

S.B.D. was supported by the National Institutes of Health/National Institute on Minority Health and Health Disparities (NIH K23 MD016230) and the Robert and Elizabeth Strickland Career Development Award, Mayo Clinic, Rochester, Minn, USA.

Declaration of Competing Interest

None of the authors has any conflict of interest.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at https://doi.org/10.1016/j.ajmo.2023.100044.

References

1. World Health Organization. World Patient Safety Day 2019. https://www.who.int/campaigns/world-patient-safety-day/2019. Accessed August 24, 2021.
2. Rodwin BA, Bilan VP, Merchant NB, et al. Rate of Preventable Mortality in Hospitalized Patients: a Systematic Review and Meta-analysis. *J Gen Intern Med*. 2020;35:2099–2106. doi:10.1007/s11606-019-05592-5.
3. National Academies of Sciences Engineering, and Medicine. *Crossing the Global Quality Chasm: Improving Health Care Worldwide*. Washington, DC: The National Academies Press; 2018.
4. Hall MJ, Levant S, DeFrances CJ. Trends in inpatient hospital deaths: National Hospital Discharge Survey, 2000-2010. Report No. 118. *NCHS Data Brief*. 2013:1–8.
5. Khera R, Haimovich J, Hurley NC, et al. Use of Machine Learning Models to Predict Death After Acute Myocardial Infarction. *JAMA Cardiol*. 2021;6:633–641. doi:10.1001/jamacardio.2021.0122.
6. Gattringer T, Posekany A, Niederkorn K, et al. Predicting Early Mortality of Acute Ischemic Stroke. *Stroke*. 2019;50:349–356. doi:10.1161/strokeaha.118.022863.
7. Angraal S, Mortazavi BJ, Gupta A, et al. Machine Learning Prediction of Mortality and Hospitalization in Heart Failure With Preserved Ejection Fraction. *JACC: Heart Fail*. 2020;8:12–21. doi:10.1016/j.jchf.2019.06.013.
8. Di MY, Liu H, Yang ZY, Bonis PA, Tang JL, Lau J. Prediction Models of Mortality in Acute Pancreatitis in Adults: A Systematic Review. *Ann Intern Med*. 2016;165:482–490. doi:10.7326/m16-0650.
9. Loke YK, Kwok CS, Niruban A, Myint PK. Value of severity scales in predicting mortality from community-acquired pneumonia: systematic review and meta-analysis. *Thorax*. 2010;65:884–890. doi:10.1136/thx.2009.134072.
10. Dugani SB, Geyer HL, Maniaci MJ, Burton MC. Perception of barriers to research among internal medicine physician hospitalists by career stage. *Hosp Pract (1995)*. 2020;48:206–212. doi:10.1080/21548331.2020.1779537.
11. Dugani SB, Geyer HL, Maniaci MJ, Fischer KM, Croghan IT, Burton C. Psychological wellness of internal medicine hospitalists during the COVID-19 pandemic. *Hosp Pract (1995)*. 2021;49:47–55. doi:10.1080/21548331.2020.1832792.
12. Dugani SB, Geyer HL, Maniaci MJ, et al. Hospitalist perspectives on barriers to recommend and potential benefit of the COVID-19 vaccine. *Hosp Pract (1995)*. 2021;49:1–7. doi:10.1080/21548331.2021.1914465.
13. Moher D, Liberati A, Tetzlaff J, Altman DGPRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Br Med J*. 2009;339:b2535. doi:10.1136/bmj.b2535.
14. Dugani S, Burton MC, Chipi P, Al-Zu'bi H, Murad MH. Risk Prediction Models for In-hospital Mortality Among General Medical Wards: A Systematic Review. PROSPERO 2020 CRD42020176054. Available at https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42020176054. Accessed August 21, 2021.
15. Covidence systematic review software, Veritas Health Innovation, Melbourne, Australia. Available at www.covidence.org. Accessed August 14, 2021.
16. Moons KGM, De Groot JAH, Bouwmeester W, et al. Critical Appraisal and Data Extraction for Systematic Reviews of Prediction Modelling Studies: The CHARMS Checklist. *PLoS Med*. 2014;11:e1001744. doi:10.1371/journal.pmed.1001744.
17. Wolff RF, Moons KGM, Riley RD, et al. PROBAST: A Tool to Assess the Risk of Bias and Applicability of Prediction Model Studies. *Ann Intern Med*. 2019;170:51. doi:10.7326/m18-1376.
18. Fabbian F, De Giorgi A, Maietti E, et al. A modified Elixhauser score for predicting in-hospital mortality in internal medicine admissions. *Eur J Intern Med*. 2017;40:37–42. doi:10.1016/j.ejim.2017.02.002.
19. Rasmussen LJH, Ladelund S, Haupt TH, Ellekilde GE, Eugen-Olsen J, Andersen O. Combining National Early Warning Score With Soluble Urokinase Plasminogen Activator Receptor (suPAR) Improves Risk Prediction in Acute Medical Patients: A Registry-Based Cohort Study. *Crit Care Medicine*. 2018;46:1961–1968. doi:10.1097/CCM.0000000000003441.
20. Moore CC, Hazard R, Saulters KJ, et al. Derivation and validation of a universal vital assessment (UVA) score: a tool for predicting mortality in adult hospitalised patients in sub-Saharan Africa. *BMJ Glob Health*. 2017;2:e000344. doi:10.1136/bmjgh-2017-000344.
21. 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 (Baltimore)*. 2017;96:e7284. doi:10.1097/MD.0000000000007284.
22. Schwartz N, Sakhnini A, Bisharat N. Predictive modeling of inpatient mortality in departments of internal medicine. *Intern Emerg Med*. 2018;13:205–211. doi:10.1007/s11739-017-1784-8.
23. Soffer S, Klang E, Barash Y, Grossman E, Zimlichman E. Predicting In-Hospital Mortality at Admission to the Medical Ward: A Big-Data Machine Learning Model. *Am J Med*. 2020;134:227–234. doi:10.1016/j.amjmed.2020.07.014.
24. Chen KF, Liu SH, Li CH, et al. Development and validation of a parsimonious and pragmatic CHARM score to predict mortality in patients with suspected sepsis. *Am J Emerg Med*. 2017;35:640–646. doi:10.1016/j.ajem.2016.10.075.
25. Fabbian F, De Giorgi A, Boari B, et al. Infections and internal medicine patients: Could a comorbidity score predict in-hospital mortality? *Medicine (Baltimore)*. 2018;97:e12818. doi:10.1097/MD.0000000000012818.
26. Gerry S, Bonnici T, Birks J, et al. Early warning scores for detecting deterioration in adult hospital patients: systematic review and critical appraisal of methodology. *Br Med J*. 2020;369:m1501. doi:10.1136/bmj.m1501.
27. Churpek MM, Yuen TC, Edelson DP. Risk Stratification of Hospitalized Patients on the Wards. *Chest*. 2013;143:1758–1765. doi:10.1378/chest.12-1605.
28. Le Gall J-R, Lemeshow S, Saulnier F. A New Simplified Acute Physiology Score (SAPS II) Based on a European/North American Multicenter Study. *J Am Med Assoc*. 1993;270:2957–2963. doi:10.1001/jama.1993.03510240069035.
29. Knaus WA, Draper EA, Wagner DP, Zimmerman JE. APACHE II: a severity of disease classification system. *Crit Care Med*. 1985;13:818–829.
30. Pirracchio R, Petersen ML, Carone M, Rigon MR, Chevret S, Van Der Laan MJ. Mortality prediction in intensive care units with the Super ICU Learner Algorithm (SICULA): a population-based study. *Lancet Respir Med*. 2015;3:42–52. doi:10.1016/s2213-2600(14)70239-5.
31. Marafino BJ, Park M, Davies JM, et al. Validation of Prediction Models for Critical Care Outcomes Using Natural Language Processing of Electronic Health Record Data. *JAMA Netw Open*. 2018;1:e185097. doi:10.1001/jamanetworkopen.2018.5097.
32. Brajer N, Cozzi B, Gao M, et al. Prospective and External Evaluation of a Machine Learning Model to Predict In-Hospital Mortality of Adults at Time of Admission. *JAMA Netw Open*. 2020;3:e1920733. doi:10.1001/jamanetworkopen.2019.20733.
33. Liu VX, Lu Y, Carey KA, et al. Comparison of Early Warning Scoring Systems for Hospitalized Patients With and Without Infection at Risk for In-Hospital Mortality and Transfer to the Intensive Care Unit. *JAMA Netw Open*. 2020;3:e205191. doi:10.1001/jamanetworkopen.2020.5191.
34. Smith MEB, Chiovaro JC, O'Neil M, et al. Early Warning System Scores for Clinical Deterioration in Hospitalized Patients: A Systematic Review. *Ann Am Thorac Soc*. 2014;11:1454–1465. doi:10.1513/annalsats.201403-102oc.
35. Churpek MM, Yuen TC, Winslow C, Meltzer DO, Kattan MW, Edelson DP. Multicenter Comparison of Machine Learning Methods and Conventional Regression for Predicting Clinical Deterioration on the Wards. *Crit Care Med*. 2016;44:368–374. doi:10.1097/ccm.0000000000001571.
36. Escobar GJ, Liu VX, Schuler A, Lawson B, Greene JD, Kipnis P. Automated Identification of Adults at Risk for In-Hospital Clinical Deterioration. *N Engl J Med*. 2020;383:1951–1960. doi:10.1056/nejmsa2001090.
37. Charlson ME, Pompei P, Ales KL, MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: Development and validation. *J Chronic Dis*. 1987;40:373–383. doi:10.1016/0021-9681(87)90171-8.
38. Elixhauser A, Steiner C, Harris DR, Coffey RM. Comorbidity Measures for Use with Administrative Data. *Med Care*. 1998;36:8–27. doi:10.1097/00005650-199801000-00004.
39. Kieszak SM, Flanders WD, Kosinski AS, Shipp CC, Karp H. A comparison of the Charlson comorbidity index derived from medical record data and administrative billing data. *J Clin Epidemiol*. 1999;52:137–142. doi:10.1016/s0895-4356(98)00154-1.
40. Van Walraven C, Dhalla IA, Bell C, et al. Derivation and validation of an index to predict early death or unplanned readmission after discharge from hospital to the community. *Can Med Assoc J*. 2010;182:551–557. doi:10.1503/cmaj.091117.
41. Van Walraven C, Forster AJ. The HOMR-Now! Model Accurately Predicts 1-Year Death Risk for Hospitalized Patients on Admission. *Am J Med*. 2017;130:991 e9-991.e16. doi:10.1016/j.amjmed.2017.03.008.
42. Schneeweiss S, Maclure M. Use of comorbidity scores for control of confounding in studies using administrative databases. *Int J Epidemiol*. 2000;29:891–898. doi:10.1093/ije/29.5.891.

43. Ou H-T, Mukherjee B, Erickson SR, Piette JD, Bagozzi RP, Balkrishnan R. Comparative Performance of Comorbidity Indices in Predicting Health Care-Related Behaviors and Outcomes among Medicaid Enrollees with Type 2 Diabetes. *Popul Health Manag*. 2012;15:220–229. doi:10.1089/pop.2011.0037.
44. Siontis GCM, Tzoulaki I, Ioannidis JPA. Predicting Death: An Empirical Evaluation of Predictive Tools for Mortality. *Arch Intern Med*. 2011;171:1721–1726. doi:10.1001/archinternmed.2011.334.
45. Riviello ED, Kiviri W, Fowler RA, et al. Predicting Mortality in Low-Income Country ICUs: The Rwanda Mortality Probability Model (R-MPM). *PLoS ONE*. 2016;11:e0155858. doi:10.1371/journal.pone.0155858.
46. Breslow MJ, Badawi O. Severity scoring in the critically ill: part 1–interpretation and accuracy of outcome prediction scoring systems. *Chest*. 2012;141:245–252. doi:10.1378/chest.11-0330.
47. Beane A, De Silva AP, De Silva N, et al. Evaluation of the feasibility and performance of early warning scores to identify patients at risk of adverse outcomes in a low-middle income country setting. *BMJ Open*. 2018;8:e019387. doi:10.1136/bmjopen-2017-019387.
48. Pouw ME, Peelen LM, Moons KGM, Kalkman CJ, Lingsma HF. Including post-discharge mortality in calculation of hospital standardised mortality ratios: retrospective analysis of hospital episode statistics. *Br Med J*. 2013;347:f5913. doi:10.1136/bmj.f5913.
49. Kristoffersen DT, Helgeland J, Clench-Aas J, Laake P, Veierød MB. Comparing hospital mortality–how to count does matter for patients hospitalized for acute myocardial infarction (AMI), stroke and hip fracture. *BMC Health Serv Res*. 2012;12:364. doi:10.1186/1472-6963-12-364.
50. Rosenthal GE, Baker DW, Norris DG, Way LE, Harper DL, Snow RJ. Relationships between in-hospital and 30-day standardized hospital mortality: implications for profiling hospitals. *Health Serv Res*. 2000;34:1449–1468.
51. Borzecki AM, Christiansen CL, Chew P, Loveland S, Rosen AK. Comparison of in-hospital versus 30-day mortality assessments for selected medical conditions. *Med Care*. 2010;48:1117–1121. doi:10.1097/MLR.0b013e3181ef9d53.
52. Churpek MM, Yuen TC, Winslow C, et al. Multicenter Development and Validation of a Risk Stratification Tool for Ward Patients. *Am J Respir Crit Care Med*. 2014;190:649–655. doi:10.1164/rccm.201406-1022oc.
53. Churpek MM, Wendlandt B, Zadravecz FJ, Adhikari R, Winslow C, Edelson DP. Association between intensive care unit transfer delay and hospital mortality: A multicenter investigation. *J Hosp Med*. 2016;11:757–762. doi:10.1002/jhm.2630.
54. Steyerberg EW, Harrell FE, Borsboom GJJM, Eijkemans MJC, Vergouwe Y, Habbema JDF. Internal validation of predictive models. *J Clin Epidemiol*. 2001;54:774–781. doi:10.1016/s0895-4356(01)00341-9.
55. Riley RD, Ensor J, Snell KIE, et al. External validation of clinical prediction models using big datasets from e-health records or IPD meta-analysis: opportunities and challenges. *Br Med J*. 2016:i3140. doi:10.1136/bmj.i3140.
56. Rudd KE, Seymour CW, Aluisio AR, et al. Association of the Quick Sequential (Sepsis-Related) Organ Failure Assessment (qSOFA) Score With Excess Hospital Mortality in Adults With Suspected Infection in Low- and Middle-Income Countries. *J Am Med Assoc*. 2018;319:2202. doi:10.1001/jama.2018.6229.
57. Schuetz P, Koller M, Christ-Crain M, et al. Predicting mortality with pneumonia severity scores: importance of model recalibration to local settings. *Epidemiol Infect*. 2008;136:1628–1637. doi:10.1017/s0950268808000435.
58. Steyerberg EW, Moons KGM, Van Der Windt DA, et al. Prognosis Research Strategy (PROGRESS) 3: Prognostic Model Research. *PLoS Med*. 2013;10:e1001381. doi:10.1371/journal.pmed.1001381.
59. Reilly BM, Evans AT. Translating Clinical Research into Clinical Practice: Impact of Using Prediction Rules To Make Decisions. *Ann Intern Med*. 2006;144:201–209. doi:10.7326/0003-4819-144-3-200602070-00009.
60. Morrison A, Polisena J, Husereau D, et al. The Effect of English-Language Restriction on Systematic Review-Based Meta-Analyses: A Systematic Review of Empirical Studies. *Int J Technol Assess Health Care*. 2012;28:138–144. doi:10.1017/s0266462312000086.
61. Dugani SB, Hydoub YM, Ayala AP, et al. Risk Factors for Premature Myocardial Infarction: A Systematic Review and Meta-analysis of 77 Studies. *Mayo Clin Proc Innov Qual Outcomes*. 2021;5:783–794. doi:10.1016/j.mayocpiqo.2021.03.009.
62. World Health Organization. The true death toll of COVID-19: Estimating global excess mortality. https://www.who.int/data/stories/the-true-death-toll-of-covid-19-estimating-global-excess-mortality. Accessed August 9, 2021.
63. Bodilsen J, Nielsen PB, Søgaard M, et al. Hospital admission and mortality rates for non-covid diseases in Denmark during covid-19 pandemic: nationwide population based cohort study. *Br Med J*. 2021;373:n1135. doi:10.1136/bmj.n1135.
64. Collins GS, Ogundimu EO, Altman DG. Sample size considerations for the external validation of a multivariable prognostic model: a resampling study. *Stat Med*. 2016;35:214–226. doi:10.1002/sim.6787.