ISSN: 3152-8686

Vol3; Issue3, Aug. 2026

JOURNAL OF TAZEEZ FOR PUBLIC HEALTH

AN OFFICIAL JOURNAL OF SAUDI HEALTH PROMOTION AND EDUCATION ASSOCIATION

Systematic review

Performance of risk scores for predicting short-term mortality in acute heart failure patients presenting to the emergency department: a systematic review

Mazi Mohammed Alanazi[1], Abrar Mohammed Albalooshy[2], Eithar Khalid Alamri[3], Mohammed Abdullah Alanazi[4], Abdulaziz Mohammed Alsahli[5], Sultan Haif Aladhyani[6], Mais Abdulrahman Joumani[7], Taif Ibrahim Alharbi[8], Muath Thaar Aladhyani[9]

1. Emergency Medicine Consultant, Head of Emergency Research Unit, Emergency Department, Riyadh First Health Cluster, Riyadh, Saudi Arabia. ORCID: 0009-0003-5836-8086.
2. Medical Intern, College of Medicine, Taibah University, Medina, Saudi Arabia. ORCID: 0009-0000-5258-5440. Email: abrar.m.albaloshi@gmail.com.
3. Medical Intern, College of Medicine, Taibah University, Medina, Saudi Arabia. ORCID: 0009-0007-9321-6503. Email: Ethar.k.alamri@gmail.com.
4. Medical Intern, College of Medicine, Northern Border University, Arar, Saudi Arabia. ORCID: 0009-0005-7857-7656. Email: mohammed.aalenezii@gmail.com.
5. Medical Intern, College of Medicine, Shaqra University, Dawadmi, Saudi Arabia. ORCID: 0009-0008-4393-2568. Email: drabdulaziz211200@gmail.com.
6. Medical Student, College of Medicine, Shaqra University, Dawadmi, Saudi Arabia. Email: sltanrooji@icloud.com.
7. Medical Intern at King Saud University and graduate of Medipol University.
8. Medical Intern, College of Medicine, Qassim University, Qassim, Saudi Arabia. ORCID: 0009-0001-6350-9088. Email: Taifharbi88@gmail.com.
9. Medical Intern, College of Medicine, Shaqra University, Dawadmi, Saudi Arabia. ORCID: 0009-0001-3382-4722. Email: mooath2022@gmail.com.

Abstract

Background: Acute heart failure (AHF) necessitate emergency department (ED) evaluation, where early mortality risk affect monitoring, admission, discharge, and follow-up decisions. Several scores have been proposed and their target outcomes, predictors, and validation settings differ. Objective: To synthesize the discrimination, calibration, and utility of risk scores predicting short-term mortality in patients presenting to EDs with AHF. Methods: Records were identified through PubMed/MEDLINE, Scopus and WOS and citation tracking, with strategies prepared for Scopus, Web of Science, and Cochrane Library through Jun 20, 2026. Eligible studies validated or

Published: 30st Aug, 2026

https://doi.org/10.65759/aeyemx64

compared multivariable scores for mortality within 30 days of emergency presentation. Study characteristics, performance and calibration findings, and risk-of-bias domains were narratively synthesized because outcome windows and populations were heterogeneous. Results: We include 19 reports. EHMRG was the most validated score for 7-day mortality, with statistics ranging from 0.73 to 0.81 in Canadian, Spanish, and United States cohorts. MEESSI-AHF showed a strong discrimination for 30-day mortality, between 0.80 and 0.84, although external cohorts identified overprediction in higher-risk groups before recalibration. EAHFE-3D show moderate discrimination for 3-day mortality, with imprecision in smaller external samples. Broader disposition scores incorporated mortality into composite outcomes and offered less direct evidence for death prediction. Conclusions: EHMRG and MEESSI-AHF have the strongest external-validation evidence. Local calibration, predictor availability, and prospective impact evaluation remain essential before disposition use.

Keywords: acute heart failure; emergency department; short-term mortality; risk score; prediction model; EHMRG; MEESSI-AHF; systematic review

Introduction

Acute heart failure (AHF) is a life-threatening syndrome characterized by rapid onset or worsening of congestion, impaired perfusion, or both, and ED assessment directs immediate treatment and site-of-care decisions (1). Heart failure admissions impose mortality, readmission, and economic burdens across health systems, with older age and multimorbidity dominating emergency populations (2). Population-based data shows that patients discharged directly from the ED retain measurable 7-day and 30-day mortality risks, and some discharged patients have risks comparable with admitted patients (3). International practice shows wide variation in admission rates, observation use, specialist involvement, and follow-up arrangements after emergency diagnosis of AHF (4).

Early risk assessment is difficult because dyspnea, edema, blood pressure, renal dysfunction, myocardial injury, and precipitating illnesses produce highly heterogeneous clinical profiles (5). A prior systematic review identified numerous emergency-focused scores with differences in predictor selection, outcome definition, modeling strategy, and validation quality (6). Subsequent clinical reviews show that disposition decisions require more than discrimination alone, because calibration, feasibility, treatment response, social support, and access to early follow-up affect safe implementation (7). These observations support structured evaluation of mortality-specific tools rather than assuming that a score predicting a composite serious outcome predicts death equally well (5,7). No single examination finding or laboratory marker provides stable mortality classification in the full emergency population, which explains the move toward multivariable models (5,6). Misclassification has asymmetric consequences, because unrecognized high risk threatens patient safety while indiscriminate admission increases crowding and resource use (3,4).

The Emergency Heart Failure Mortality Risk Grade was designed for 7-day mortality and combines readily available presentation variables, selected comorbidities, laboratory results, prehospital transport, and outpatient metolazone use (6). The Multiple Estimation of Risk Based on the ED Spanish Score in Patients with AHF targets 30-day mortality through a broader model that includes clinical status, laboratory findings, electrocardiographic information, and functional indicators (6). The EAHFE-3D scale addresses death within 3 days, whereas the Ottawa Heart Failure Risk Scale, STRATIFY, and HEARTRISK6 estimate serious composite outcomes that include death with escalation of care or rehospitalization (5–7).

This systematic review aimed to synthesize the performance of multivariable risk scores for short-term mortality in adults presenting to ED with AHF. The primary outcomes were discrimination and calibration for death within 3, 7, or 30 days, with secondary attention to threshold performance, risk-group separation, external validity, missing predictors, and clinical usability.

Methods

Design and reporting framework

This review was structured according to PRISMA 2020 principles for systematic reviews and framed as a prognostic-model evidence synthesis (Fig 1). The review question followed a population, index model, outcome, and setting framework. The population comprised adults presenting to an ED with AHF, the index models were multivariable risk scores available during emergency assessment, and the outcome was all-cause short-term mortality.

Information sources and search strategy

PubMed/MEDLINE, Scopus and WOS was searched from inception through Jun 20, 2026, and equivalent database-specific strategies were prepared for Scopus, Web of Science Core Collection, and the Cochrane Library. Search concepts combined controlled vocabulary and free-text terms for acute heart failure, emergency department, mortality, prognosis, prediction model, clinical score, validation, discrimination, calibration, EHMRG, MEESSI-AHF, and EAHFE-3D. Reference lists of eligible reports and relevant systematic reviews were screened for additional studies. Searches were restricted neither by language nor publication year, and translated abstracts were planned for non-English records. Search results are intended to be exported with complete citation metadata and de-duplicated before screening.

The complete electronic strategy for PubMed was: (“acute heart failure” OR “decompensated heart failure”) AND (“emergency department” OR emergency) AND (mortality OR death) AND (“risk score” OR “prediction model” OR prognos* OR validation).

Eligibility and study selection

Eligible reports enrolled adults aged 18 years or older with AHF evaluated in an ED and derived, internally validated, externally validated, updated, or compared a multivariable score. Studies had to report mortality within 30 days of ED presentation, in-hospital mortality originating from an ED cohort, or a composite outcome with death reported as a distinct component. We exclude reviews, editorials, conference abstracts without adequate data, single-biomarker models, inpatient-only cohorts lacking emergency presentation data, and studies focused solely on long-term outcomes. Two reviewers screened titles, abstracts, and full texts independently, with disagreements resolved through discussion and a third reviewer.

30/08/2026

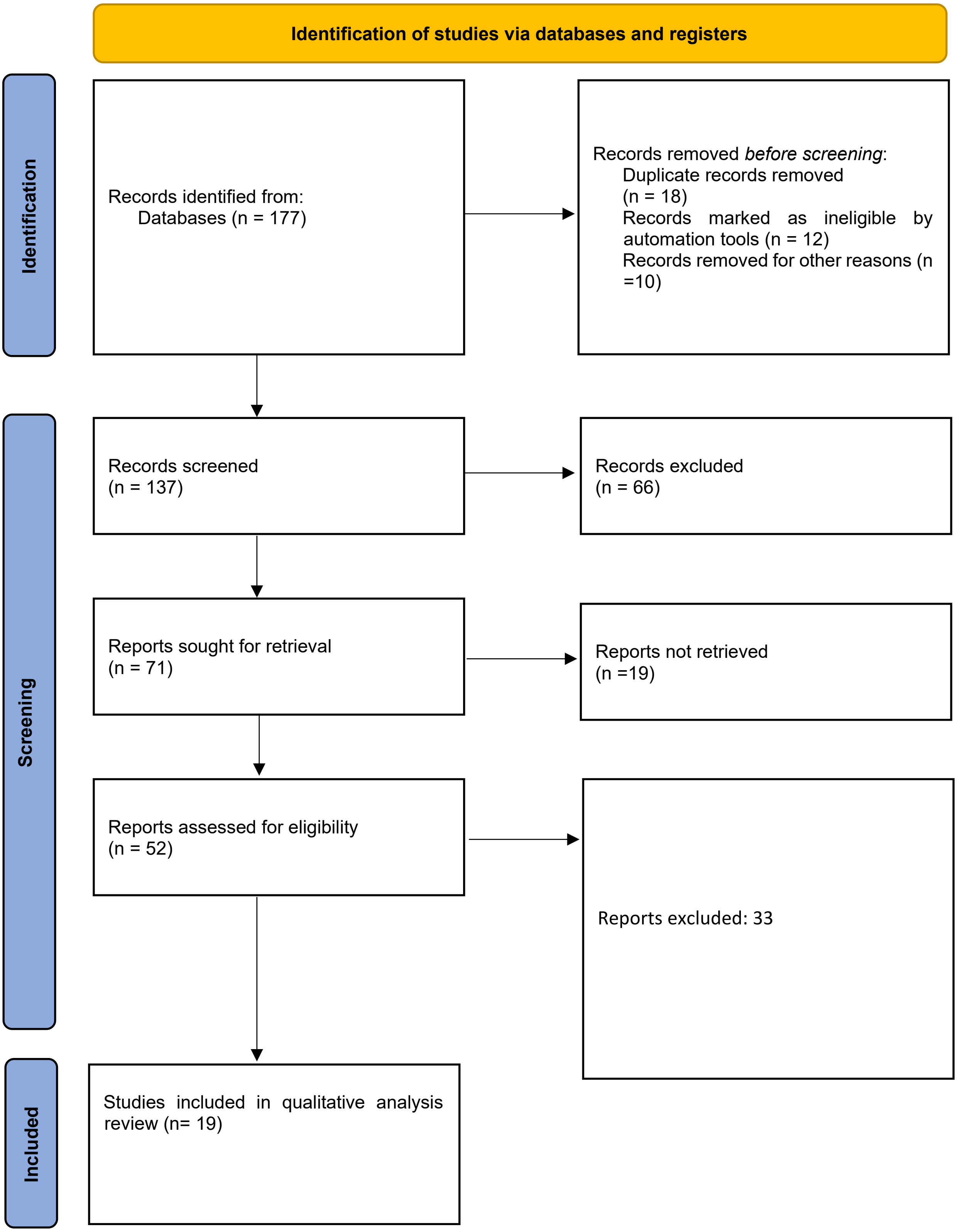
ISSN: 3152-8686

Data extraction, quality assessment, and synthesis

The extraction form included country, design, sample size, score, prediction horizon, outcome incidence, discrimination, calibration, threshold measures, missing-data handling, and validation type. Risk of bias and applicability were assessed with the four PROBAST domains: participants, predictors, outcome, and analysis. Meta-analysis was not undertaken because prediction horizons, score versions, outcome definitions, and reporting metrics differed substantially. Results were qualitatively analyzed, with priority given to external validation, calibration, and performance in clinically relevant low-risk strata.

ISSN: 3152-8686

Fig 1: PRISMA consort chart



Results

Study profile and principal findings

We include 19 original reports published from 2008 through 2026, including mortality-specific models and broader emergency disposition scores in which death formed part of a serious-outcome composite (6). The Acute Heart Failure Index validation included 8,384 admitted patients and classified 19.2% as low risk; 30-day mortality in that subgroup was 2.9% (8). The original EHMRG cohort included 12,591 emergency patients, and the model produced c-statistics of 0.806 in derivation and 0.804 in validation for 7-day mortality (9). The Ottawa Heart Failure Risk Scale was derived in 559 emergency visits for a composite serious adverse event, whereas STRATIFY later targeted 30-day adverse events rather than isolated mortality (10,11).

EAHFE-3D was developed from 6,597 patients in 34 Spanish EDs and showed areas under the curve near 0.80 in derivation and 0.76 in validation for 3-day mortality (12). The Acute Heart Failure Risk Score was created in three Spanish EDs for poor short-term clinical course, using variables available during early assessment (13). MEESSI-AHF used 4,867 derivation and 3,229 validation patients and achieved c-statistics of 0.836 and 0.828, respectively, for 30-day mortality with strong reported internal calibration (14). Prospective validation of the Ottawa scale confirmed sensitivity for serious outcomes, although death is combined with nonfatal events and quantitative natriuretic peptide testing affected operational performance (15).

In a Spanish cohort of 1,553 patients, EHMRG discrimination for 7-day mortality was 0.74, and recalibration improved agreement between predicted and observed risk (16). External validation of EAHFE-3D included 717 patients with only seven deaths (17). The prospective Canadian ACUTE study enrolled 1,983 patients and reported a c-statistic of 0.81 for EHMRG 7-day mortality (18). Swiss validation of MEESSI-AHF found a c-statistic of 0.80 for 30-day mortality and identified overprediction in higher-risk strata before intercept recalibration (19). A population-based Alberta cohort of 6,708 ambulance-transported patients reported EHMRG c-indices of 0.73 for 7-day mortality and 0.71 for 30-day mortality (20).

MEESSI-AHF retained strong discrimination for mortality measured from emergency arrival, including c-statistics of 0.85 at 7 days and 0.83 at 30 days, while performance fell for revisit and admission composites (21). A multicenter United States EHMRG study found 7-day mortality near 2% and a modest improvement in discrimination when EHMRG was added to a baseline model (22). Conventional logistic EHMRG modeling performed comparably with several machine-learning approaches in 12,608 patients (23). A later United States retrospective validation reinforced risk-group separation while identifying workflow and missing-variable limitations in routine records (24). HEARTRISK6 was derived from three prospective Canadian cohorts totaling 2,246 patients and included 70 deaths within a broader short-term serious-outcome endpoint (25). A 2026 Turkish prospective validation of 237 patients reported EHMRG areas under the curve of 0.741 for 7-day mortality and 0.755 for 30-day mortality, with no deaths in the two lowest categories and negative predictive values above 97% (26). In mortality-specific tools, EHMRG had the broadest external-validation base for 7-day death, while MEESSI-AHF showed the highest and most consistent discrimination for 30-day death (9,14,16,18–24,26). Calibration drift was recurrent when scores moved between countries, institutions, transport patterns, and outcome incidences (16,19,20,26). EAHFE-3D offered rapid 3-day estimation, although fewer deaths and smaller external samples produced greater statistical uncertainty than the principal EHMRG and MEESSI-AHF validations (12,17).

The principal derivation studies used large multicenter cohorts, although retrospective data collection, predictor selection, and internal validation created some analytical concerns (9,12–14). External validations reduced development-related optimism, while several studies had limited event counts, complete-case analysis, or incomplete calibration reporting (16,17,19,22,24,26). Applicability was strongest for EHMRG because its variables are routinely collected, whereas MEESSI-AHF and several composite scores require functional, reassessment, or biomarker data that are absent in some emergency workflows (7,14,15,19,25). The provisional PROBAST assessment therefore favored the major prospective external validations and rated small-event or retrospective single-system studies less favorably (17–20,22,24,26).

Table 1. Characteristics of the original studies included in the evidence synthesis

Study	Country; design	Sample	Score	Outcome	Principal finding
Hsieh et al., 2008 (8)	United States; retrospective validation	8,384	Acute Heart Failure Index	In-hospital complications and 30-day death	19.2% were classified as low risk; 30-day mortality in that group was 2.9%.
Lee et al., 2012 (9)	Canada; population-based derivation and validation	12,591	EHMRG	7-day mortality	C-statistic 0.806 in derivation and 0.804 in validation.
Stiell et al., 2013 (10)	Canada; prospective derivation	559	OHFRS	30-day serious adverse event composite	Designed for disposition-related serious outcomes rather than isolated mortality.
Collins et al., 2015 (11)	United States; prospective derivation with internal validation	1,033	STRATIFY	30-day adverse event composite	Identified a lower-risk subgroup; mortality-specific calibration was not the principal target.
Jacob et al., 2016 (12)	Spain; multicenter derivation and validation	6,597	EAHFE-3D	3-day mortality	AUC 0.80 in derivation and 0.76 in validation.

30/08/2026

Study	Country; design	Sample	Score	Outcome	Principal finding
Garcia-Gutierrez et al., 2017 (13)	Spain; prospective derivation and validation	1,430	AHFRS	Poor short-term clinical course	AUC 0.83 in derivation and 0.82 in validation for the composite endpoint.
Miró et al., 2017 (14)	Spain; multicenter derivation and validation	8,096	MEESSI-AHF	30-day mortality	C-statistic 0.836 in derivation and 0.828 in validation; strong internal calibration.
Stiell et al., 2017 (15)	Canada; prospective external validation	1,100	OHFRS	30-day serious adverse event composite	Confirmed sensitivity for serious outcomes; mortality represented one component.
Gil et al., 2018 (16)	Spain; external validation	1,553	EHMRG	7-day mortality	C-statistic 0.741; recalibration improved model fit.
Garcia-Gutierrez et al., 2018 (17)	Spain; external validation	717	EAHFE-3D	3-day mortality	AUC 0.76; only seven deaths produced wide uncertainty.
Lee et al., 2019 (18)	Canada; prospective multicenter validation	1,983	EHMRG7 and EHMRG30-ST	7-day and 30-day mortality	C-statistics 0.81 and 0.77; EHMRG7 exceeded physician risk estimation.
Wussler et al., 2019 (19)	Switzerland; prospective external validation	1,572 enrolled; 1,247 complete cases	MEESSI-AHF	30-day mortality	C-statistic 0.80; high-risk over-prediction improved after intercept recalibration.
Sepehrvand et al., 2019 (20)	Canada; population-based external validation	6,708	EHMRG	7-day and 30-day mortality	C-indices 0.73 and 0.71, with clear separation across risk categories.
Rossello et al., 2021 (21)	Spain; secondary cohort analysis	4,867	MEESSI-AHF	Multiple short-term outcomes	C-statistics 0.85 for 7-day and 0.83 for 30-day mortality; weaker for revisit composites.
Garg et al., 2021 (22)	United States; retrospective	3,315	EHMRG	7-day mortality	Adding EHMRG increased the c-statistic from 0.748 to 0.772.

Study	Country; design	Sample	Score	Outcome	Principal finding
	multicenter validation				
Austin et al., 2022 (23)	Canada; comparative modeling study	12,608	EHMRG logistic model versus machine learning	7-day and 30-day mortality	Conventional regression performed comparably and showed stronger calibration than several machine-learning models.
Nelson et al., 2023 (24)	United States; retrospective single-system validation	304	EHMRG	7-day and 30-day mortality	No 7-day deaths occurred in very-low- and low-risk groups; sample size limited precision.
Stiell et al., 2024 (25)	Canada; analysis of three prospective cohorts	2,246	HEARTRISK6	Short-term serious outcome composite	281 serious outcomes included 70 deaths; evidence chiefly supports disposition safety.
Ünlü et al., 2026 (26)	Türkiye; prospective external validation	237	EHMRG	7-day and 30-day mortality	AUC 0.741 and 0.755; no deaths in the two lowest categories and NPV above 97%.

Table 2. Comparative interpretation of the principal score families

Score family	Intended prediction target	Performance in reviewed studies	Main strengths	Principal limitations
EHMRG	7-day mortality; 30-day extension available	C-statistics or AUCs generally 0.73–0.81 for 7-day mortality	Broadest international validation; routinely collected predictors; clear risk groups	Calibration drift in settings; retrospective validation in several cohorts; metolazone and transport variables reduce transportability
MEESSI-AHF	30-day mortality	C-statistics generally 0.80–0.84, with 0.85 reported for 7-day death in secondary analysis	Strong discrimination; large multicenter derivation; detailed risk separation	More predictors; complete-case analysis in external validation; high-risk overprediction before recalibration

Score family	Intended prediction target	Performance in reviewed studies	Main strengths	Principal limitations
EAHFE-3D	3-day mortality	AUC 0.76–0.80	Focused on very early death; emergency-specific derivation	Few events in external validation; wide confidence intervals; limited calibration evidence
Acute Heart Failure Index	In-hospital complications and 30-day mortality among admitted patients	Low-risk 30-day mortality 2.9%	Simple low-risk classification	Admitted-only population; limited applicability to direct emergency discharge
OHFRS, STRATIFY, AHFRS, and HEARTRISK6	Serious adverse event or poor-course composites	Variable threshold performance; not directly comparable with mortality-only AUCs	Incorporate deterioration, escalation of care, and revisit-related safety	Composite endpoints mix fatal and nonfatal events; mortality-specific calibration is incomplete
Overall evidence	Short-term emergency prognosis	Best-supported tools are EHMRG for 7-day death and MEESSI-AHF for 30-day death	Multiple large derivation and external-validation cohorts	Few impact trials; uneven geographic representation; inconsistent missing-data and calibration reporting

30/08/2026

ISSN: 3152-8686

Discussion

This review identifies EHMRG and MEESSI-AHF as the mortality scores with the most convincing ED validation evidence (6,9,14,16,18–24,26). EHMRG repeatedly separated very-low-risk from high-risk patients for 7-day death, with discrimination ranging from moderate to good across Canadian, Spanish, United States, and Turkish settings (9,16,18,20,22,24,26). MEESSI-AHF achieved stronger discrimination for 30-day death (14,19,21). The practical trade-off is greater data demand, because full MEESSI-AHF calculation depends on variables that are not uniformly available at initial triage or in resource-limited departments (7,14,19).

The external-validation findings demonstrate why discrimination alone is insufficient for selecting a score (16,19,20,26). A model with an acceptable c-statistic produces misleading absolute risks when baseline mortality, referral patterns, ambulance use, laboratory assays, treatment pathways, or discharge practices differ from the derivation setting (27). The observed overprediction of high-risk MEESSI-AHF patients and improved EHMRG fit after recalibration indicate that local intercept adjustment deserves formal evaluation before disposition thresholds are adopted (16,19). PROBAST principles also place substantial weight on representative participants, prespecified predictors, adequate event numbers, missing-data methods, and avoidance of overfitting, domains that were incompletely reported in several retrospective validations (28). Consistent TRIPOD reporting would improve assessment of model equations, calibration plots, threshold selection, and reproducibility (29).

Prediction horizon affected apparent performance, with stronger discrimination generally observed for outcomes nearer the emergency presentation (6,12,17,21). EHMRG answers a different clinical question from MEESSI-AHF: the former estimates imminent 7-day death, whereas the latter estimates a broader 30-day period that incorporates post-discharge vulnerability (9,14). EAHFE-3D focuses on early death, although its smaller number of external events limits confidence in precise thresholds (12,17). Ottawa, STRATIFY, AHFRS, and HEARTRISK6 offer useful disposition frameworks, although composite serious outcomes combine events of unequal clinical importance and obscure mortality-specific calibration (10,11,13,15,25). A score with high sensitivity for a composite endpoint is not automatically an accurate mortality model (5,6).

Clinical implementation requires a separation between prognosis and disposition authority (7). Very-low-risk categories with high negative predictive value support consideration of discharge only after confirmation of hemodynamic stability, treatment response, renal status, oxygenation, self-care capacity, medication access, and rapid follow-up (18,26). High-risk classification supports closer monitoring and senior review, although score-guided admission has not been established through randomized impact trials in the principal mortality-score literature (18,25). Shared decision-making remains important because patient preferences, frailty, cognitive status, caregiver support (4,7). Electronic calculation integrated into triage or the medical record would reduce arithmetic burden, although automation requires safeguards for missing values, unit errors, and outdated coefficients (27,29).

ISSN: 3152-8686

The review has several limitations. Direct pooling was inappropriate because studies used 3-day, 7-day, 30-day, in-hospital, and composite outcomes, with different score versions and validation methods. Several cohorts were retrospective, some external validations contained few deaths, and calibration reporting was less consistent than discrimination reporting. The studies were concentrated in Canada, Spain, Switzerland, and the United States, with limited representation from Africa, the Middle East, and low-resource emergency systems.

Conclusion

EHMRG is best supported for 7-day mortality and offers practical risk-group separation using routinely collected variables. MEESSI-AHF provides stronger 30-day discrimination, with greater predictor requirements and recurrent need for external recalibration. EAHFE-3D addresses 3-day death, although external estimates remain less precise. Composite disposition scores supply complementary safety information rather than interchangeable mortality estimates. EDs adopting any score require local validation, calibration assessment, complete predictor capture, and integration with treatment response, frailty, social circumstances, follow-up access, and clinician judgment before disposition decisions.

Abbreviations

AHF, acute heart failure; AHFRS, Acute Heart Failure Risk Score; AUC, area under the receiver operating characteristic curve; ED, emergency department; EAHFE-3D, Epidemiology of Acute Heart Failure in Emergency Departments–3-Day scale; EHMRG, Emergency Heart Failure Mortality Risk Grade; EHMRG30-ST, 30-day EHMRG extension incorporating ST-segment depression; MEESSI-AHF, Multiple Estimation of Risk Based on the Emergency Department Spanish Score in Patients With Acute Heart Failure; NIV, noninvasive ventilation; NPV, negative predictive value; NT-proBNP, N-terminal pro-B-type natriuretic peptide; OHFRS, Ottawa Heart Failure Risk Scale; PRISMA, Preferred Reporting Items for Systematic Reviews and Meta-Analyses; PROBAST, Prediction Model Risk of Bias Assessment Tool; SAE, serious adverse event; TRIPOD, Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis.

References

1. McDonagh TA, Metra M, Adamo M, et al. 2021 ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure. Eur Heart J. 2021;42(36):3599–3726. doi:10.1093/eurheartj/ehab368.
2. Ambrosy AP, Fonarow GC, Butler J, et al. The global health and economic burden of hospitalizations for heart failure: lessons learned from hospitalized heart failure registries. J Am Coll Cardiol. 2014;63(12):1123–1133.
3. Lee DS, Schull MJ, Alter DA, et al. Early deaths in patients with heart failure discharged from the emergency department: a population-based analysis. Circ Heart Fail. 2010;3(2):228–235. doi:10.1161/CIRCHEARTFAILURE.109.885285.

4. Miró Ò, Levy PD, Möckel M, et al. Disposition of emergency department patients diagnosed with acute heart failure: an international emergency medicine perspective. Eur J Emerg Med. 2017;24(1):2–12.
5. Michaud AM, Parker SIA, Ganshorn H, et al. Prediction of early adverse events in emergency department patients with acute heart failure: a systematic review. Canadian Journal of Cardiology. 2018;34(2):168–179.
6. Miró Ò, Rossello X, Gil V, et al. Risk stratification scores for patients with acute heart failure in the emergency department: a systematic review. Eur Heart J Acute Cardiovasc Care. 2020;9(5):375–398. doi:10.1177/2048872620930889.
7. Fountoulaki K, Ventoulis I, Dourakis E, et al. Emergency department risk assessment and disposition of acute heart failure patients: existing evidence and ongoing challenges. Heart Fail Rev. 2023;28(4):781–793. doi:10.1007/s10741-022-10272-4.
8. Hsieh M, Auble TE, Yealy DM. Validation of the Acute Heart Failure Index. Ann Emerg Med. 2008;51(1):37–44. doi:10.1016/j.annemergmed.2007.07.026.
9. Lee DS, Stitt A, Austin PC, et al. Prediction of heart failure mortality in emergent care: a cohort study. Ann Intern Med. 2012;156(11):767–775.
10. Stiell IG, Clement CM, Brison RJ, et al. A risk scoring system to identify emergency department patients with heart failure at high risk for serious adverse events. Acad Emerg Med. 2013;20(1):17–26.
11. Collins SP, Jenkins CA, Harrell FE Jr, 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. doi:10.1016/j.jchf.2015.05.007.
12. Jacob J, Miró Ò, Herrero P, 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.
13. Garcia-Gutierrez S, Quintana JM, Antón-Ladislao A, et al. Creation and validation of the acute heart failure risk score: AHFRS. Intern Emerg Med. 2017;12(8):1197–1206. doi:10.1007/s11739-016-1541-4.
14. Miró Ò, Rossello X, Gil V, 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. doi:10.7326/M16-2726.
15. Stiell IG, Perry JJ, Clement CM, et al. Prospective and explicit clinical validation of the Ottawa Heart Failure Risk Scale, with and without use of quantitative NT-proBNP. Acad Emerg Med. 2017;24(3):316–327. doi:10.1111/acem.13141.
16. Gil V, Miró Ò, Schull MJ, 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. Eur J Emerg Med. 2018;25(3):169–177.
17. García-Gutiérrez S, Quintana López JM, Antón-Ladislao A, et al. External validity of a prognostic score for acute heart failure based on the Epidemiology of Acute Heart Failure in Emergency Departments Registry:

the EAHFE-3D scale. Emergencias. 2018;30(2):84–90.

18. Lee DS, Lee JS, Schull MJ, et al. Prospective validation of the Emergency Heart Failure Mortality Risk Grade for acute heart failure. Circulation. 2019;139(9):1146–1156. doi:10.1161/CIRCULATIONAHA.118.035509.
19. Wussler D, Kozhuharov N, Sabti Z, et al. External validation of the MEESSI Acute Heart Failure Risk Score: a cohort study. Ann Intern Med. 2019;170(4):248–256. doi:10.7326/M18-1967.
20. Sepehrvand N, Youngson E, Bakal JA, et al. External validation and refinement of Emergency Heart Failure Mortality Risk Grade risk model in patients with heart failure in the emergency department. CJC Open. 2019;1(3):123–130. doi:10.1016/j.cjco.2019.03.003.
21. Rossello X, Bueno H, Gil V, et al. MEESSI-AHF risk score performance to predict multiple post-index event and post-discharge short-term outcomes. Eur Heart J Acute Cardiovasc Care. 2021;10(2):142–152.
22. Garg N, Pekmezaris R, Stevens G, et al. Performance of Emergency Heart Failure Mortality Risk Grade in the emergency department. West J Emerg Med. 2021;22(3):672–677. doi:10.5811/westjem.2021.1.48978.
23. Austin DE, Lee DS, Wang CX, et al. Comparison of machine learning and the regression-based EHMRG model for predicting early mortality in acute heart failure. Int J Cardiol. 2022;365:78–84. doi:10.1016/j.ijcard.2022.07.035.
24. Nelson DM, Madsen BE, Kopecky SL, et al. Retrospective validation of acute heart failure risk stratification in the emergency department. Heart Lung. 2023;57:31–40. doi:10.1016/j.hrtlng.2022.08.005.
25. Stiell IG, Perry JJ, Clement CM, et al. The HEARTRISK6 Scale: predicting short-term serious outcomes for emergency department patients with acute heart failure. JACC Adv. 2024;3(7):100988. doi:10.1016/j.jacadv.2024.100988.
26. Ünlü B, Demir O, Özpolat Ç, et al. External prognostic validation of the EHMRG score for risk stratification and disposition decisions in acute heart failure. BMC Emerg Med. 2026;26(1):145. doi:10.1186/s12873-026-01573-0.
27. Van Calster B, McLernon DJ, van Smeden M, Wynants L, Steyerberg EW. Calibration: the Achilles heel of predictive analytics. BMC Med. 2019;17(1):230. doi:10.1186/s12916-019-1466-7.
28. Moons KGM, Wolff RF, Riley RD, et al. PROBAST: a tool to assess risk of bias and applicability of prediction model studies: explanation and elaboration. Ann Intern Med. 2019;170(1):W1–W33. doi:10.7326/M18-1377.
29. Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis: the TRIPOD Statement. Ann Intern Med. 2015;162(1):55–63. doi:10.7326/M14-0697.