Occhino *et al. Systematic Reviews* (2026) 15:99
https://doi.org/10.1186/s13643-026-03100-5

Systematic Reviews

RESEARCH

Open Access

Prognostic models in populations with heart failure: a systematic review and meta-analysis

Giuseppe Occhino[1,2,3], Alessandro Musa[2], Anita Andreano[4], Pietro Magnoni[4], Martino Bussa[1], Deborah Testa[4], Adele Zanfino[4], Matteo Petrosino[1,2], Maria Grazia Valsecchi[1,3], Lucia Bisceglia[2], Paola Rebora[1,3*], Antonio Giampiero Russo[4] and the PROPHET-I. study group

Abstract

Background Heart failure (HF) remains a major cause of morbidity and mortality, highlighting the need for reliable prognostic models. This study provides a systematic review and meta-analysis of prognostic models focused on mortality, hospitalization, and their composite event.

Methods We screened 2271 papers and reviewed 58 prognostic models from 44 studies involving 362,759 HF patients. The predictive performance of these models was assessed, and a meta-analysis was performed for the Seattle Heart Failure Model (SHFM), which focuses on mortality outcomes at 1 year. The models were evaluated via the PROBAST tool for risk of bias and applicability.

Results Of the 58 models, 86% underwent internal and/or external validation in independent cohorts, with statistical models (88%) being more common than machine learning approaches (12%). Clinical data were used in 79% of the models, whereas the remaining models used electronic health records (EHR) or mixed sources of data. Mortality models ($n=40$) revealed a 1-year discrimination range between 0.66 and 0.89. The most common predictors included age, renal function, blood pressure, coronary artery disease, and serum sodium. A meta-analysis of 5 studies that applied the SHFM at 1 year revealed a pooled *C*-statistic of 0.71 (95% CI: 0.64–0.78), with relatively low heterogeneity ($\tau^2=0.003$). Hospitalization models ($n=9$) demonstrated discrimination up to 0.86, and composite event models ($n=9$) showed similar predictive power. The risk of bias was high in 88% of the models, largely due to univariable predictor selection and handling/reporting of missing values.

Conclusions This systematic review highlighted the heterogeneity of HF prognostic models and patient populations in terms of severity and symptoms, emphasizing challenges in developing commonly applicable tools. Most studies enrolled patients with reduced ejection fraction (EF), whereas evidence for HF with preserved EF was limited. Despite widespread research, few HF prognostic models meet current standards for clinical implementation. The large majority of the studies did not report calibration and had a poor alignment with contemporary therapies. Future model development should prioritize transparency, methodological rigor, and external validation.

Systematic review registration PROSPERO CRD42023488017.

Keywords Heart failure, Prognostic models, Mortality prediction, Meta-analysis, Clinical predictors

*Correspondence:
Paola Rebora
paola.rebora@unimib.it
Full list of author information is available at the end of the article

Background

Heart failure (HF) is a clinical condition characterized by symptoms and/or signs resulting from a structural or functional problem in the heart, confirmed by elevated natriuretic peptide levels or objective evidence of increased fluid in the lungs or body [1]. This broad clinical definition encompasses multiple etiologies and phenotypes, which are commonly distinguished based on left ventricular ejection fraction (LVEF) preservation. HF affects 1% to 2% of adults worldwide, with the prevalence increasing to 10% or more in people aged 70 and older [2]. In the first year following a diagnosis of HF, the death rate in the USA and Europe varies between 10 and 40% [3]. As the global population ages, its prevalence is rising, leading to an increasing burden of disease on individuals and healthcare systems.

The development of strategic tools aimed at planning health promotion/prevention interventions at the population level has the potential to reduce the burden of HF [4]. In this context, the use of automated calculators fitting predictive models on routinely collected electronic health records could be very useful [5]. For example, periodically estimating the risk of progression for each HF patient would be useful for tailoring secondary and tertiary prevention and adapting the setting of care and follow-up. In order to develop such tools, a comprehensive review of known prognostic factors and models would be very useful. Among known prognostic factors of HF progression, we can identify the New York Heart Association (NYHA) class, LVEF, hypertension, diabetes, obesity, kidney failure, liver failure, smoking, and B-type natriuretic peptide [6]. They have been included in various predictive models. The Seattle Heart Failure Model (SHFM) [7] was developed to predict 1-, 2-, and 3-year survival in heart failure patients and has been validated in several other cohorts. The MAGGIC heart failure score was developed to predict mortality in patients from 30 cohort studies [8]. It comprises demographic data and laboratory values as well as data on device therapy and medication. Several other prognostic models have been developed for the prediction of heart failure progression [9], including both statistical models [8, 10–12] and, more recently, applications of machine learning approaches [13–16]. Both of these classes of models use health administrative databases and/or specific medical databases as sources of data and can include many different demographic, socioeconomic, and clinical prognostic factors.

The available systematic reviews on the prognosis of heart failure patients [11, 13, 14, 16–18] revealed great variability in the available risk prediction models with respect to patient population and modelling and poor reporting quality. Additionally, the most updated reviews focused exclusively on machine learning methods. In general, little attention has been given to the heterogeneity of HF patients. Moreover, no meta-analysis of predictive performances has been done on specific prognostic models, such as the Seattle Heart Failure Model.

The aim of the present systematic review was to obtain an updated view of the available prognostic models for heart failure patients. In particular, this review considers all studies that involve adult human subjects with heart failure (population); it focuses on multivariable regression models and machine learning techniques for predicting heart failure progression on the basis of demographic, socioeconomic, and clinical prognostic factors (models). A secondary objective is to perform a quantitative synthesis of the predictive performance of specific prognostic models whenever enough studies are available and to identify their sources of heterogeneity. This includes models developed for different populations using different risk factors and methodologies. By doing so, we aim to better understand the factors contributing to differences across studies and gain a more comprehensive view of how these models perform in different populations. The clinical outcomes considered in this systematic review were overall mortality, HF-related and all-cause hospitalization, emergency department access, and stroke as a major adverse cardiovascular event (MACE). We considered predictions from 1 to 5 years since enrollment. These outcomes are the most frequently examined outcomes in predictive models for HF because they reflect critical aspects of disease management and patient health.

Methods

The study was performed within the framework of an EU-funded project (PNRR-MAD-2022–1237603). It was conducted and reported following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses of Diagnostic Test Accuracy Studies (PRISMA-DTA) guidelines (checklist in Additional file 1) [19]. The protocol was registered on PROSPERO (PROSPERO 2023 CRD42023488017; available from https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42023488017).

The review question and the adopted inclusion criteria are described via the PICOTS framework [20] as follows:

- Population: This review considered all studies that involved adult human subjects with heart failure, without a defined ethnicity. The definition of adult age for each primary study was accepted. All definitions from primary studies were accepted, considering that the recent universal definition was published in 2021 [1].

- Index Model: Multivariable regression models and machine learning techniques for predicting heart failure progression on the basis of demographic, socioeconomic, and clinical prognostic factors. We included studies that developed multivariable prognostic models (using either regression-based or machine learning approaches) as well as studies that performed external validation of existing models.
- Outcome: Predictive performance of the model for predicting overall mortality, HF-related and all-cause hospitalizations, emergency department access, and stroke as a major adverse cardiovascular event (MACE) within 5 years of inclusion. All performance measures were considered for the systematic review, but meta-analysis was feasible only on the *C*-statistic, a discrimination measure that quantifies the model's ability to correctly rank pairs of individuals who did and did not experience the event.
- Timing: from 1 to 5 years.
- Setting: Prediction models intended to be used by healthcare professionals at any time during heart failure progression.

A priori, the models for the different outcomes were described separately, and the predictive performance of the models with the same predictors was quantitatively summarized.

Both cohort and case–control studies, as well as observational studies incorporating cohorts from randomized controlled trials (RCTs), were assessed for inclusion. Cross-sectional studies were excluded.

Search methods for the identification of studies

The search strategy was designed to access both the PubMed and EMBASE full-text archives and considered all the articles regarding predictive and prognostic models published between January 1st, 2012, and November 15th, 2023, with no language restrictions applied. Non-English articles were eligible, and translation support (including professional interpreters, if needed) was planned to enable screening and data extraction. The search string used combinations of terms related to HF, prediction/prognosis, modeling methods, and disease progression or outcomes. It was examined and validated by a librarian, and then, it was applied to the specified digital archives. The full strings are available in the Additional file 2. Duplicated records were removed via automatic procedures and subsequently manually checked. The deduplicated library of references was imported into CADIMA [21]. In addition to this search strategy, we manually checked the reference lists of the studies included as full texts and of previous systematic reviews of prognostic models for any heart failure complications. Abstracts and proceedings were evaluated for inclusion in the review. Letters and commentaries were not included in the review, but they were used as supporting material for the primary studies.

Selection process

During the screening process, titles and abstracts were independently assessed by two reviewers (GO, AM). Eligible records were assessed in full text for eligibility criteria. Disagreements were resolved by a third author (AA). After that, we proceeded to critical appraisal and data extraction via an ad hoc developed and piloted data extraction form tested on three studies. To develop the form, we used the Checklist for critical Appraisal and Data Extraction for Systematic Reviews of Prediction Modeling Studies (CHARMS) [22].

Data collection

Regarding general information, the following items were recorded: first author, title, journal, year of publication, study design during the phase of development of the model and during the phase of validation (cohort from either observational study or RCT, nested case–control, case–cohort), start and end of recruitment, mean or median age (expressed in years, with standard deviation or interquartile range, depending on how it was reported), sex (% male), geographical location, and time with heart failure (expressed in years). Moreover, we recorded whether the study involved only some subgroups of HF patients (NYHA class, ejection fraction value, etiology of HF) and the patient eligibility criteria (inclusion and exclusion).

We extracted the specific endpoints predicted by the model (mortality, hospitalization, emergency department access, and stroke), the proportion of events at the last follow-up (%), the type of endpoint (e.g., single or combined endpoints), the time of follow-up, and whether the endpoint was assessed without knowledge of the candidate predictors.

We included studies reporting time-to-event outcomes with a minimum follow-up of 1 year. For each study, we recorded the type of endpoint and whether the statistical analysis accounted for the timing of events (e.g., Cox proportional hazards models). While some studies appropriately modelled time-to-event data, others reported only cumulative event rates without accounting for the timing of events. Failure to appropriately account for time-to-event data was considered a methodological limitation and contributed to a higher risk of bias in the analysis domain of the PROBAST assessment. This information was extracted to ensure transparency and to guide interpretation of model performance.

We extracted the number of different models, data sources (clinical, administrative, and mixed), if the predictors were blinded to the outcome and the timing of predictor measurement (e.g., at patient presentation, at diagnosis, at treatment initiation, and active call of the cohort extracted from the database), the sample size, the modelling method (logistic, survival, neural network, or machine learning techniques), the use of shrinkage, the methods/criteria for selecting predictors along with the resulting list, and, lastly, for multivariable regression models, the final equation. When more models were reported on the same outcome and on the same sample, the one with the best performance was used.

Concerning multivariable prediction models, discrimination (i.e., the ability of the model to distinguish between participants who experience the outcome and those who do not) is commonly quantified by the *C*-statistic or the Area Under the Receiver Operating Characteristic (ROC) Curve (AUC). The *C*-statistic quantifies the probability that a randomly selected individual who experiences the event has a higher predicted risk than a randomly selected individual who does not experience the event. Calibration refers to the agreement between predicted risks and observed outcomes, and can be assessed for example through calibration slope or observed-to-expected ratios. The Brier score is an overall performance metric capturing calibration and discrimination aspects. With respect to machine learning approaches, the most commonly used predictive performance measure is accuracy, which expresses the ratio between the number of correct predictions and the total number of predictions. We compared multivariable prediction models and machine learning models through the *C*-statistic or AUC measure, which provides an aggregate measure of performance across all possible methods. We extracted data at five different possible times (1 to 5 years). Finally, class imbalance and the handling of missing values were reported.

When confidence intervals of the performance measures were not available, we estimated them by using the Hanley–McNeil Wald method [23] given the reported sample size and number of events.

The developed data extraction form was subjected to an initial testing phase before implementation in CADIMA. The data were extracted by two different authors, who resolved any disagreements by discussion. If a disagreement persisted, a third author resolved the disagreement. If data were missing, we contacted the corresponding author to obtain the information.

Risk of bias assessment

Risk of bias (RoB) was assessed using the *prediction model risk of bias assessment tool* (PROBAST) [24], designed for prognostic model studies. It comprises 4 domains and includes a total of 20 signaling questions that aid in the assessment of risk of bias (i.e., the likelihood that study design or analysis flaws distort the estimated predictive performance) and concerns regarding applicability. The participants domain covers potential sources of bias and applicability related to the data source used and how the participants were selected (appropriate study design data source and inclusion and exclusion criteria of the participants). The second domain relates to the predictors and addresses possible biases in relation to how the predictors are defined and measured. In the support for judgment section, reviewers had the option to list and explain how the predictors were defined, when they were assessed, and whether additional information was available at the time of assessment. The third domain covered the definition and determination of the outcome. We analyzed if the outcome was determined appropriately, if there was a prespecified or standard outcome definition used, if any predictors were excluded from the outcome definition, if the outcome was defined in a similar way for all participants, if the outcome was determined without knowledge of predictor information, and if the time interval between predictor assessment and outcome determination was appropriate. The last domain assessed the correct handling of essential statistical factors. It examined whether key statistical considerations were correctly addressed with questions about sample size, model building, handling missing data, etc. Internal (i.e., assessment of model performance within the development dataset) and external validation (i.e., evaluation in an independent population) were included as items in the quality evaluation. The questions within PROBAST were responded to with options such as "yes" (Y), "probably yes" (PY), "no" (N), "probably no" (PN), or "no information" (NI). Each signaling question was formulated such that a "yes" response suggests a low risk of bias (RoB), whereas a "no" response indicates a high risk of bias. The inclusion of "probably yes" (PY) and "probably no" (PN) allowed assessments to be made when there was insufficient information to confidently answer with a "yes" or "no". If relevant information was missing for some of the signaling questions, there was an unclear risk of bias. When a prediction model evaluation received low ratings in all domains, it was deemed suitable to assign an overall judgment of "low ROB" or "low concern regarding applicability". However, if the evaluation received a high rating in at least one domain, it was considered "high RoB" or "high concern regarding applicability". In cases where the evaluation was unclear in one or more domains but had low ratings in the remaining domains, it was categorized as having an "unclear RoB" or "unclear concern regarding applicability".

Data synthesis

Once the data were collected and the quality of the studies was assessed, we explored the number of models predicting each outcome at each time point by the same set of potential features (i.e., same index prognostic model) [25]. We then performed a meta-analysis only on those index prognostic models that had been validated in at least five studies (with the same time horizon). This threshold was chosen because it ensures robust statistical power, enhances the representativeness of results, minimizes publication bias, and allows for a comprehensive examination of methodological diversity and effect variability [26]. Consequently, index models that were not validated in at least 5 studies were not included in the meta-analysis. Since we anticipated heterogeneity due to population/case mix, study design, study setting, and differences in management approaches and measurement methods, we conducted a meta-analysis via a random-effects approach. We stratified the meta-analysis by different time points of outcomes. The effect measure used for the meta-analysis was logit-transformed *C*-statistic. To evaluate statistical heterogeneity, we used the between-study variance Tau ^{2} [27]. With a sufficient number of studies (more than 10 studies), the performance of predictive models for predefined subgroups (sex, age group: adult vs. elderly, HF stage/NYHA class) and the causes of heterogeneity would have been further explored through subgroup analyses. A sensitivity analysis excluding studies with a high risk of bias was planned with a sufficient number of studies available.

Results

Literature search

We identified 2271 potentially relevant articles from electronic databases reporting results on prognostic models for HF (PRISMA flow diagram —Fig. 1). Of these, 2149 were excluded after title and abstract screening. The full texts of 2 articles were not accessible. After full-text assessment of the remaining 120 studies, 76 were excluded for the following reasons: not respecting the inclusion criteria for the target population ($n=14$), not including multivariable predictive models ($n=16$), and not respecting the inclusion criteria for the outcome measure ($n=28$) or timing ($n=7$).

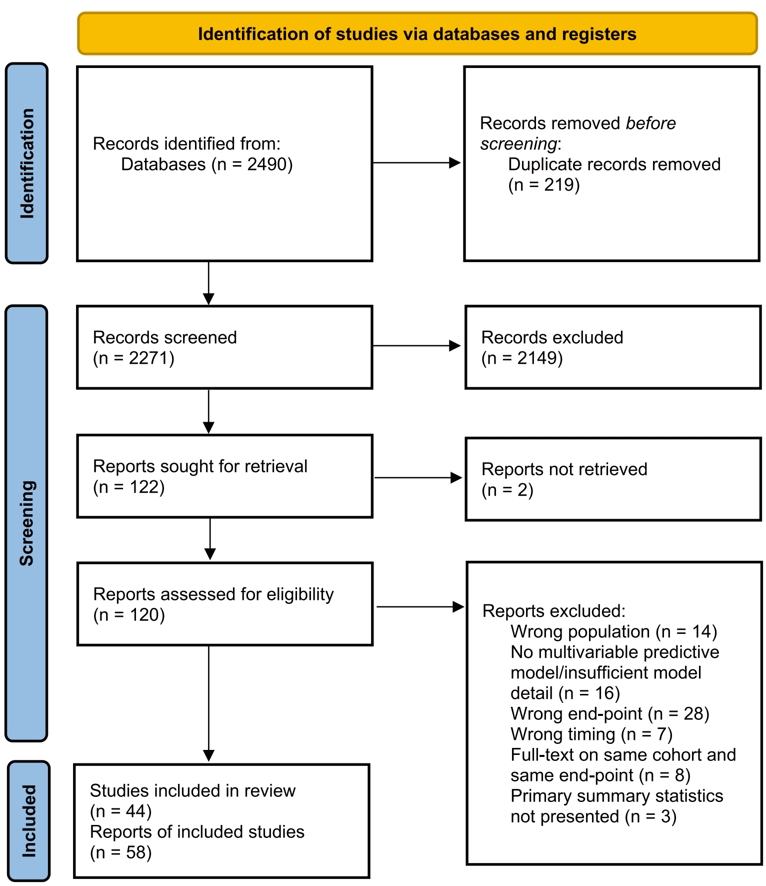

Fig. 1 PRISMA flow diagram for the identification of the studies

Additionally, 8 studies were duplicate publications on the same cohort and same endpoint as the other included studies. Primary summary statistics were not presented in 3 studies.

The remaining 44 articles were included in the systematic review, which included a total of 58 prognostic models. All studies included in the final analysis were published in English, and all screened articles provided an English abstract.

Study characteristics

The 44 eligible studies (Table 1) included a total of 362,759 patients with HF. Eight studies were based in the USA, 4 in Spain, 4 in China, 3 in Italy, 3 in the UK, 2 in Japan, and 2 in Singapore, whereas the other studies were mainly based in European or Asian countries or included different countries.

Most of the studies ($n=34$, 77%) were based on clinical data (26 from observational cohort studies and 8 from randomized trials), whereas 8 used data from electronic health data and 2 from mixed sources. The mean age ranged between 53 and 78 years, and the percentage of males ranged between 39 and 98%. The median follow-up among studies ranged between 1 and 5 years, with 10 studies presenting a follow-up of 1 year and 8 with a follow-up of at least 4 years.

Almost all the studies ($n=42$) reported left-sided HF, whereas 2 did not report the type of HF. Most of the studies ($n=35$, 80%) did not include a selection criterion on the basis of the NYHA class, whereas 5 included only patients with an NYHA class between II and IV, 3 included patients with an NYHA class between III and IV, and 1 included only patients with class II/III. With respect to LVEF, 23 studies did not use any inclusion criteria on the basis of this criterion, while 14 included only patients with reduced EF (HFrEF, from < 30% to < 40%), 3 included patients with mildly reduced (HFmrEF) or reduced EF (up to < 50%), and the remaining 4 focused on patients with preserved EF (HFpEF, > 50%, $n=2$) or patients with mildly reduced or preserved EF (up to > 40%, $n=2$). With respect to etiology, only 1 study included specifically patients with chronic HF of ischemic cause, whereas the others did not report the causes of HF (27, 61%) and/or included different causes. Given the heterogeneity of the patient population in Table 1, when possible, we classified the studies according to the mean value of EF [2] and to the universal definition of HF [1], and we used these values to separate the studies that focused on patients with (i) advanced HF ($n=12$), (ii) symptomatic or not with R + mR-reduced EF ($n=7$), (iii) symptomatic with preserved EF ($n=7$) or studies that included any stage of HF ($n=18$). Among the latter, only one study [54] stratified patients according to EF.

Model characteristics

Fifty (86%) of the 58 models underwent internal and/or external validation, but calibration was very poorly reported: only 4 reported the calibration slope, and 3 the Brier score. Almost all the models were statistical-based approaches ($n=51$, 88%), with 45 models applying the Cox proportional hazards model. Interestingly, almost all machine learning-based approaches used methods based on decision trees or their extensions (3 decision trees, 2 random forests, and one XGBoost). Only one used dynamic radius means (extensions of the fixed radius nearest neighbor algorithm) after feature selection based on the orthogonal relief algorithm. Among the 51 statistical models, 28 did not apply any selection method in the multivariable model, whereas 10 applied backward selection of predictors, 9 applied stepwise or forward selection, and the remaining 4 performed both backward and forward selection.

The most common outcome was all-cause mortality (40 out of 58), followed by hospitalization (9 out of 58) and the composite outcome of death or hospitalization (8 out of 58). Only one study considered stroke as the endpoint, in combination with mortality and myocardial infarction [38].

Risk of bias and applicability

Figure 2 shows the distribution of the overall risk of bias across the 58 predictive models within each of the four PROBAST domains: participants, predictors, outcomes, and analysis (details in Additional file 3). The models generally reported a low risk of bias for predictors and outcome assessment. In the participants' domain, 62% of the models presented a low risk of bias, whereas 21% presented a high risk, mainly due to the exclusion of patients with missing data, with 17% marked as unclear. For the predictor domain, 86% of the models were rated as low risk, and 14% were rated as high risk. In the outcome domain, almost all the models (91%) presented a low risk of bias, whereas 7% presented a high risk, and 2% were classified as unclear. The analysis domain was the one with the highest number of models presenting a high risk of bias ($n=43$, 74%), mainly due to the selection of predictors on the basis of univariable analysis ($n=32$, 55%) and inadequate handling/reporting of missing data ($n=10$, 17%). On the other hand, 21% of the models ($n=12$) were low risk, and 5% ($n=3$) were unclear.

With respect to the overall judgment of risk of bias, 6 models (10%) were classified as low risk, while 51 (88%) had high risk, and only one (2%) had an unclear risk of bias. Considering applicability, the overall judgment rated most of the models ($n=46$, 79%) as low concern, 10 (17%) as high concern, and only 2 (4%) as unclear concern.

Table 1 Characteristics of the studies included in the systematic review

Study reference	Country/continent	Study design/data source	Data collection	Patient population	Sample size	Age (SD)	Male %	EF (SD)	Universal definition of HF	Classification based on EF	Outcome(s) assessed
Advanced HF											
Oh, C. (2012) [28]	South Korea	Cohort–clinical	2007–2009	Patients with NYHA class II/IV, LVEF ≤ 35% with a QRS interval > 120 ms	66	66 (11)	70	26 (7)	C-D	R	Death
Regoli, F. (2013) [29]	Europe	Cohort–clinical	2002–2011	Patients who underwent CRT device implantation	1139	67.2 (10.7)	77.4	25 (8.2)	C-D	R	Death
Richter, B. (2013) [30]	Austria	Cohort–clinical	2003–2004	Current hospitalization due to cardiac decompensation, NYHA class III/IV, LVEF < 40% and/or cardiothoracic ratio > 0.5	349	73.3 (14.1)	66.2	19.4 (NR)	C-D	R + mR	Death
Scrutinio, D. (2014) [31]	Italy	Cohort–clinical	2005–2011	Hospitalization for worsening of CHF, history of HF, NYHA Class III/IV, LVEF ≤ 30%, and need for intravenous diuretic and/or inotropic treatment	445	62 (13)	85	22.9 (4.6)	C-D	R	Death
Smith, T. (2012) [32]	Netherlands and Switzerland	Cohort–clinical	2000–2009	Patients with symptomatic HF, LVEF ≤ 35%, QRS duration ≥ 120 ms	413	61 (12)	76	24 (7)	C-D	R	Death
Ketchum, E.S. (2012) [33]	North America/Europe	Randomized trial	2005–2008	NYHA II/III HF patients with EF ≤ 35%	961	62 (12)	80	27 (6)	C	R	Death
O'Connor, C.M. (2012) [34]	North America and France	Randomized trial	2003–2007	Outpatients with LVEF < 35% and NYHA class II/IV, receiving ACE-inhibitor and beta-adrenergic blockade for ≥ 6 weeks	2331	59.3 (12.6)	72	25 (7.4)	C-D	R	Death, death or hospitalization
Simpson, J. (2020) [35]	World	Randomized trial	2009–2014	Patients ≥ 18 years, NYHA class II-IV, LVEF ≤ 35%, BNP ≥ 150 pg/mL or NTproBNP ≥ 600 pg/mL, taking ACE inhibitor, ARB, beta blocker or mineralocorticoid receptor antagonist	7016	63.2 (11.7)	78.3	28.4 (5.7)	C-D	R	Death

Table 1 (continued)

Study reference	Country/continent	Study design/data source	Data collection	Patient population	Sample size	Age (SD)	Male %	EF (SD)	Universal definition of HF	Classification based on EF	Outcome(s) assessed
Upshaw, J.N. (2016) [36]	North America/Oceania	Randomized trial	1997–2001	Age ≥ 18 years with NYHA Class II or III chronic, stable HF with LVEF ≤ 35%	2521	59 (12)	77	25 (7.4)	C	R	Death, hospitalization
Zhang, J. (2013) [37]	Europe	Randomized trial	2000–2002	Hospitalization for worsening HF; LVEF < 40%, LV end-diastolic dimension > 30 mm/m; receiving furosemide ≥ 40 mg/day or equivalent	284	68.7 (10.4)	78	24.7 (7.5)	C-D	R	Death, death or hospitalization
Nymo, S.H. (2014) [38]	Europe/Africa	Randomized trial	2003–2005	Age > 60 years with chronic HF of ischemic cause, NYHA class II/IV, and with LVEF ≤ 40% (≤ 35% if NYHA II)	1464	71.77 (6.87)	76.7	32 (7)	C-D	R	Death, death or hospitalization, death or stroke or MI
Li, H. (2020) [39]	China	Cohort–clinical	2014–2019	Age of ≥ 18 years; diagnosis of HFrEF; EF < 40% or 40–50% with structural heart disease or diastolic dysfunction; NYHA class II–IV	878	69.3 (10.7)	66.6	Not reported	C-D	R + mR	Death
Symptomatic or not with HFrEF + HFmrEF											
Clemens, M. (2012) [40]	Hungary	Cohort–clinical	2004–2010	Patients referred for CRT	427	61.6 (11.1)	73	26.9 (6.2)	B-D	R + mR	Death
Vishram-Nielsen, J.K.K. (2020) [41]	Canada	Cohort–clinical	2000–2012	Adults ≥ 18 years of age with a diagnosis of HF with reduced LVEF (≤ 40%)	1136	55.4 (13.2)	75.3	26 (8)	B-D	R	Death
Laszczynska, O. (2016) [42]	Portugal	Cohort–clinical	2000–2011	Patients with HF based on the European Society of Cardiology, and history of depressed left ventricular systolic function (LVSF)	565	68.3 (15.6)	68.5	Not reported	B-D	R + mR	Death
Bayes-Genis, A. (2014) [43]	Spain	Cohort–clinical	2006–2010	At least 1 HF hospitalization, or a reduced LVEF	876	69.3 (12.4)	71.6	34.3 (12.6)	B-D	R + mR	Death, death or hospitalization

Table 1 (continued)

Study reference	Country/continent	Study design/data source	Data collection	Patient population	Sample size	Age (SD)	Male %	EF (SD)	Universal definition of HF	Classification based on EF	Outcome(s) assessed
Fontanive, P. (2013) [44]	Italy	Cohort–clinical	2002–2010	Patients with history of chronic HF, LVEF ≤ 45% and etiology due to ischemic or hypertensive heart disease or idiopathic cardiomyopathy	388	69 (12)	77	32 (9)	B-D	R + mR	Death
Siriopol, D. (2021) [45]	Romania	Cohort–clinical	2016–2018	Patients with LVEF ≤ 45%	151	67.1 (12.1)	69.5	32.5 (10.2)	B-D	R + mR	Death
Panahiazar, M. (2015) [46]	United States	Cohort–EHR	1993–2013	EF ≤ 50% within 2 months of HF diagnoses; no prior diagnosis of CAD, myocarditis, infiltrative cardiomyopathy, and severe valvular disease	5044	78 (10)	52	36 (10.3)	B-D	R + mR	Death
Symptomatic with HFpEF											
Barlera, S. (2013) [47]	Italy	Randomized trial	2002–2005	Age ≥ 18 years, chronic symptomatic HF, EF > 40%, hospital admitted at least once in the preceding year for HF	6975	67.2 (10.3)	78.3	33.4 (8.4)	C-D	mR + P	Death
Han, Y. (2023) [48]	China	Cohort–clinical	2016–2018	Age ≥ 18 years, and hospitalized with a primary diagnosis of new-onset HF or decompensation of chronic HF, HFpEF patients	1442	69 (12.3)	50.5	59.6 (8.2)	C-D	P	Death
Kanagala, P. (2020) [49]	United Kingdom	Cohort–clinical	Not reported	LVEF > 50% on TTE and age ≥ 18 years	140	73 (9)	49	56 (5)	B-D	P	Death or hospitalization
Kasahara, S. (2019) [50]	Japan	Cohort–clinical	2006–2010	HF patients with LVEF ≥ 50%	170	70.8 (11)	51	60 (9)	C-D	P	Death
Mendez Fernandez, A.B. (2020) [51]	Spain	Cohort–clinical	2010–2015	Patients after an episode of decompensated HF, either after an emergency room visit for dyspnea or hospital admission for HF	311	72 (13)	56	58 (12)	C	mR + P	Death
Sun, Y. (2021) [52]	China	Cohort–clinical	2015–2018	Patients hospitalized with HFpEF	476	70.5 (8.4)	39.3	56.7 (2.7)	C-D	P	Death, hospitalization

Table 1 (continued)

Study reference	Country/continent	Study design/data source	Data collection	Patient population	Sample size	Age (SD)	Male %	EF (SD)	Universal definition of HF	Classification based on EF	Outcome(s) assessed
Angraal, S. (2020) [53]	America	Randomized trial	2006–2013	patients aged ≥ 50 years with at least 1 sign and at least 1 symptom of HF and LVEF ≥ 45% to receive spironolactone or placebo therapy	1088	71.7 (11.1)	50	58.3 (8.2)	C-D	P	Death, hospitalization
Any											
Nadruz, W. (2017) [54]	United States	Cohort–clinical	2007–2012	Undergone clinically indicated CPET	969	55 (14)	73	33.4 (7.5)	B-D	R/mR/P	Hospitalization
Hammadah, M. (2014) [55]	United States	Cohort–clinical	2001–2006	Age ≥ 18 years, plasma samples available for Cp measurements	890	66.6 (10)	61	36.7 (18.6)	B-D	R + mR + P	Death
Honold, J. (2013) [56]	Germany	Cohort–clinical	2002–2008	Patients with chronic IHD who received intracoronary progenitor cell application presented in a HF outpatient clinic for CPET testing	157	60 (12)	86.6	39 (11)	B-C	R + mR + P	Death
Jackson, C.E. (2015) [57]	United Kingdom	Cohort–clinical	2006–2009	Age ≥ 18 years and BNP > 100 pg/ml	628	70.8 (10.6)	58	40.1 (12.1)	B-D	R + mR + P	Death
Kadowaki, S. (2016) [58]	Japan	Cohort–clinical	2009–2010	Patients admitted for evaluation or treatment of CHF	134	71 (13)	60	48 (19)	B-D	R + mR + P	Death or hospitalization
Ky, B. (2012) [59]	United States	Cohort–clinical	2003–2012	Patients with HF severity ranging from mild disease to severe disease requiring advanced therapies	1513	56 (15)	66	34 (17)	B-D	R + mR + P	Death
Voors, A.A. (2017) [60]	United Kingdom	Cohort–clinical	2010–2014	Patients aged > 18 years with a HF diagnosis based on echocardiographic evidence of left ventricular (LV) dysfunction or a previous admission with HF treated with furosemide ≥ 20 mg/day or equivalent	1738	73.7 (10.7)	65.9	41 (13)	B-D	R + mR + P	Death, hospitalization, death or hospitalization

Table 1 (continued)

Study reference	Country/continent	Study design/data source	Data collection	Patient population	Sample size	Age (SD)	Male %	EF (SD)	Universal definition of HF	Classification based on EF	Outcome(s) assessed
Zafrir, B. (2012) [61]	Israel	Cohort–clinical	2005–2009	HF patients, both with reduced and preserved systolic function	500	67 (13)	70	34.8 (16.1)	B-D	R + mR + P	Death
Escobar, A. (2017) [62]	Spain	Cohort–clinical	2009–2013	Patients with primary diagnosis of HF, being over 18 years of age	1282	77.4 (10.1)	50.7	50.8 (16.3)	B-D	R + mR + P	Death
Williams, B.A. (2018) [63]	United States	Cohort–EHR	2003–2014	Patients received healthcare services for at least 2 years; and having a pre-existing or new diagnosis of HF defined by ICD-9 codes as a primary or secondary diagnosis at either 1 inpatient or 2 outpatient encounters	26,851	73 (13)	50	50 (15)	B-D	R + mR + P	Death
Bowen, G.S. (2018) [64]	United States	Cohort–mixed	2013–2015	Patients with HF defined as a primary diagnostic code of HF on admission	30,111	71.1 (11.2)	98.1	40.4 (16.2)	B-D	R + mR + P	Death
Bulluck, H. (2019) [65]	Singapore	Cohort–EHR	2008–2013	Patients presenting to hospital with a STEMI within 12 h of symptoms onset and were reperfused by PPCI	2280	58.3 (11.2)	84.5	45 (14.8)	B-D	R + mR + P	Hospitalization
Jing, L. (2020) [66]	United States	Cohort–EHR	2001–2019	Patients with diagnosis of HF	548	73.3 (14.1)	57	45.3 (18.6)	B-D	R + mR + P	Death
Sartipy, U. (2014) [67]	Sweden	Cohort–EHR	2000–2012	Patients with clinician-judged heart failure	51,043	75.1 (12.1)	60	40.6 (21.7)	B-D	R + mR + P	Death
Yap, J. (2019) [68]	Singapore	Cohort–EHR	2008–2009	Patients ≥ 21 years and admitted with the DRG code 252 (HF)	729	71.4 (11.6)	43.5	52.9 (30)	B-D	R + mR + P	Death
Quiros-Lopez, R. (2019) [69]	Spain	Cohort–EHR	2008–2018	Patients aged ≥ 50 years and having a diagnosis of HF according to the European Society of Cardiology	3337	81	Not reported	Not reported	B-D	R + mR + P	Death

Table 1 (continued)

Study reference	Country/continent	Study design/data source	Data collection	Patient population	Sample size	Age (SD)	Male %	EF (SD)	Universal definition of HF	Classification based on EF	Outcome(s) assessed
Wang, L. (2012) [70]	United States	Cohort–mixed	2008–2009	All patients with ≥ 1 primary or secondary diagnosis of HF that occurred 1 year before the index date of June 1, 2009	198,640	73(NR)	98.1	Not reported	B-D	R + mR + P	Death, hospitalization, death or hospitalization
Wang, Z. (2018) [71]	China	Cohort–EHR	2009–2016	Patients with ICD-10-CM diagnosis of HF; with a minimum of 1 HF therapy initiated within the first 2 days of hospitalization	4353	Not reported	Not reported	Not reported	B-D	R + mR + P	Death

SD standard deviation, *HFpEF* heart failure with preserved ejection fraction, *HFrEF* heart failure with reduced ejection fraction, *HFmrEF* heart failure with mid-range ejection fraction

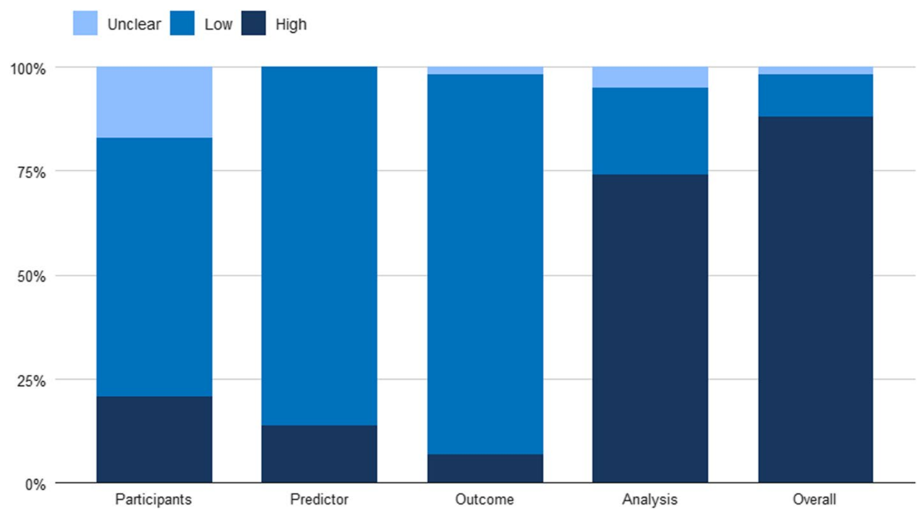

Fig. 2 Risk of bias of the included studies

Mortality

Among the 40 models for all-cause mortality, the percentage of events at the last follow-up ranged between 7 and 56%. A summary of these models is reported in Table 2. Discrimination is reported by the *C*-statistic, where higher values indicate better ability to separate patients with and without events.

Twenty-one validated studies reported discrimination at 1 year, ranging between 0.66 (95% CI: 0.62–0.70) and 0.89 (95% CI: 0.85–0.93), with 8 studies using data from EHRs ($n=6$) or mixed sources ($n=2$). Fourteen studies reported discrimination at 2 years, 8 at 3 years, 4 at 4 years, and 7 at 5 years. Calibration has been very poorly described, with only a few studies reporting Brier scores or calibration slopes.

We counted the number of times each predictor was included in the abovementioned models. The ten predictors that recurred most frequently were as follows: age, renal function, blood pressure, presence of coronary artery disease (CAD), sodium value, NYHA class, ejection fraction, sex, weight or body mass index (BMI), and B-type natriuretic peptide (Table 3).

Twenty-six out of the 40 models did not use a previously derived score but developed their own score, of which two did not validate it. Among the other models, most used the Seattle Heart Failure Model, sometimes improved with additional predictors, whereas other predictive models (such as the MAGGIC, ADHF/NT-proBNP, or H2FPEF score) did not appear more than once as the best performing model in our systematic review. Considering the 6 studies reporting the predictive performance of the SHFM at any time point, we performed a meta-analysis of the 5 studies presenting the *C*-statistic at 1 year (Fig. 3), since at the following time points, fewer than 5 studies reported the discrimination of the SHFM. The pooled *C*-statistic was 0.71 (95% CI: 0.64–0.78). The estimated 95% CI prediction interval was 0.50–0.86, and the random effects model showed low heterogeneity across studies, with a between-study variance of $\tau^2=0.003$. Higgins' and Thompson's I^2 values were high ($I^2=90.2\%$) because of the high number of included patients [27]. In fact, when the larger study by Williams et al. [63] was removed, the I^2 was 80.1%, but the pooled C-statistic remained similar (0.73, 95% CI: 0.64–0.81). As expected, all five models used statistical-based approaches and mainly regarded symptomatic HF patients with reduced ejection fraction, with only one study also including patients with preserved EF. Four out of the five models presented low concern regarding applicability, whereas only one [41] presented a low risk of bias. The main concern regarded the lack of reporting of inclusion criteria [29, 32], the definition and assessment of predictors [32, 63], and the handling of missing data [42, 63]. A sensitivity analysis excluding studies with risk of bias was not performed because of the low number of studies.

Hospitalization

With respect to hospitalization, we found 9 models (derived from 7 studies, Table 4), of which 2 regarded all-cause hospitalizations (with 1 and 3 years of follow-up) and the others regarded HF-related hospitalizations, with the percentage of events ranging from 5.4% in the study of Bulluck et al. [65] with 1 year of follow-up to 65% in the study of Sun et al. [14] with a median follow-up of 2.3 years. Five studies out of 7 used clinical data, whereas the other two used data from EHRs ($n=1$) or mixed sources ($n=1$).

Table 2 Discrimination performance and main characteristics of the models on mortality

Study reference	Modelling method	Variables included in the final model	Use of previously developed risk scores	N. event/sample size	Median follow-up (years)	1-year *C*-statistic	2-year *C*-statistic	3-year *C*-statistic	5-year *C*-statistic	Validation (internal or external)
Advanced HF										
Oh, C. (2012) [28]	Cox model	Prior stroke, HR, sodium, creatinine	No	13/66	1.9		0.8 (0.66–0.94)			Yes
Regoli, F. (2013) [29]	Logistic Regression	Gender, ischemic HF etiology, weight, CRT-P device, cholesterol	SHFM	300/1139	3.3	0.66 (0.62–0.7)	0.67 (0.63–0.71)		0.68 (0.64–0.72)	Yes
Richter, B. (2013) [30]	Cox model	Age, sTRAIL, NT-proBNP, sFAS, GDF-15, fractalkine, HGF, COPD	No	195/349	4.9				0.81 (0.76–0.85)	Yes
Scrutinio, D. (2014) [31]	Cox model	COPD, SBP, eGFR, sodium, Hb, NT-proBNP, LVEF, moderate-to-severe tricuspid regurgitation, prior hospitalization	ADHF/NT-proBNP	144/445	1	0.75 (0.7–0.8)				Yes
Smith, T. (2012) [32]	Cox model	Age, gender, weight, SBP, NYHA class, LVEF, cause of left ventricular dysfunction, ACEi, ARB, β-blocker, statin, daily diuretic dose, sodium, digoxin, carvedilol, creatinine, Hb	SHFM	78/413	2.8	0.78 (0.72–0.84)	0.75 (0.69–0.81)	0.7 (0.64–0.76)	0.7 (0.64–0.76)	Yes

Table 2 (continued)

Study reference	Modelling method	Variables included in the final model	Use of previously developed risk scores	N. event/ sample size	Median follow-up (years)	1-year *C*-statistic	2-year *C*-statistic	3-year *C*-statistic	5-year *C*-statistic	Validation (internal or external)
Ketchum, E.S. (2012) [33]	Cox model	Age, gender, SBP, ischemic origin, NYHA class, EF, sodium, creatinine, ACEi, ARB, β-blocker, carvedilol, statin, digoxin, furosemide, ICD benefit, MIBG H/M	SHFM (improved)	101/961	2	0.73 (0.68–0.78)				No
O'Connor, C.M. (2012) [34]	Cox model	Gender, exercise duration on CPX test, BMI, creatinine, mitral regurgitation, LVEF, urea, ventricular conduction prior to CPX, DBP, Canadian Angina Classification, treatment	No, development of HF-ACTION	387/2331	2.5		0.74 (0.71–0.76)			Yes
Simpson, J. (2020) [35]	Cox model	Age, gender, race/ethnicity, HF duration, NYHA class, LVEF, diabetes, MI, PAD, β-blocker, allocation to sacubitril/valsartan, prior percutaneous coronary intervention, BMI, HR, SBP, HHF, lipid lowering therapy	No	1587/7016	3.1	0.71 (0.69–0.74)	0.7 (0.68–0.72)	0.69 (0.68–0.7)		Yes
Upshaw, J.N. (2016) [36]	Cox model	Age, gender, NYHA class, LVEF, creatinine, sodium, SBP, weight, DM, IHD, AF, stroke	No	666/2521	3.8			0.71 (0.69–0.73) *		Yes

Table 2 (continued)

Study reference	Modelling method	Variables included in the final model	Use of previously developed risk scores	N. event/ sample size	Median follow-up (years)	1-year C-statistic	2-year C-statistic	3-year C-statistic	5-year C-statistic	Validation (internal or external)
Zhang, J. (2013) [37]	Decision Tree	Age, NT-proBNP, sinus rhythm, SBP, sodium, MI, randomization cohort	No	69/284	1	0.89 (0.85–0.93)				Yes
Nymo, S.H. (2014) [38]	Cox model	NT-proBNP, IL-8	No	329/1464	2.7			0.7 (0.67–0.73)		No
Li, H. (2020) [39]	Cox model	Age, NYHA class, N-T pro-BNP, renal dysfunction, LVMI, percutaneous coronary intervention, AF	No	195/878	3			0.8 (0.78–0.83)		Yes
Symptomatic or not with HFrEF + HFmrEF										
Clemens, M. (2012) [40]	Logistic Regression	Age, gender, NYHA class, EF, weight, SBP, etiology of cardiomyopathy, QRS width, pharmacologic, device, laboratory parameters, LV end-diastolic diameter, LV lead	SHFM (improved)	72/427	2.7	0.81 (0.73–0.88)	0.82 (0.76–0.88)		0.78 (0.67–0.88)	Yes
Vishram-Nielsen, J.K.K. (2020) [41]	Cox model	Age, gender, weight, SBP, ischemic cardiomyopathy, NYHA class, LVEF, dialysis, CRT-P, CRT-D, ICD, Hb, leucocytes, lymphocyte, creatinine, UA, sodium, cholesterol, ACE/ARBs, β-blockers, statins	SHFM	231/1136	3	0.76 (0.72–0.81)		0.73 (0.69–0.76)		Yes

Table 2 (continued)

Study reference	Modelling method	Variables included in the final model	Use of previously developed risk scores	N. event/ sample size	Median follow-up (years)	1-year C-statistic	2-year C-statistic	3-year C-statistic	5-year C-statistic	Validation (internal or external)
Laszczynska, O. (2016) [42]	Cox model	Age, gender, NYHA class, weight, SBP, LVEF, ischemic cause, ACEi, ARB, β-blocker, statin, daily diuretic dose, implanted devices, Hb, lymphocytes, UA, cholesterol, sodium	SHFM	245/565	5	0.72 (0.65–0.78)		0.71 (0.66–0.77)	0.69 (0.63–0.75)	Yes
Bayes-Genis, A. (2014) [43]	Cox model	Age, gender, LVEF, eGFR, NYHA class, DM, ischemic etiology, Hb, sodium, β-blocker, ACEi, ARB, NT-proBNP, ST2	No	392/876	4.2				0.77 (0.75–0.79)	No
Fontanive, P. (2013) [44]	Cox model	Age, NYHA class, LVEF, eGFR, NT-proBNP, ACEi, b-blockers, furosemide daily dose, plasma levels, TAPSE, LVend-diastolic volume index, moderate to-severe MR	No, development of Clinical and Echocardiographic Score	99/388	2.4		0.77 (0.72–0.83)			Yes
Siriopol, D. (2021) [45]	Decision Tree	NYHA class, sodium, urea, SBP	No	53/151	1.7		0.79 (0.72–0.86)			Yes

Table 2 (continued)

Study reference	Modelling method	Variables included in the final model	Use of previously developed risk scores	N. event/ sample size	Median follow-up (years)	1-year C-statistic	2-year C-statistic	3-year C-statistic	5-year C-statistic	Validation (internal or external)
Panahiazar, M. (2015)^ [46]	Cox model	Age, gender, EF, ischemic disease, SBP, ACEi, ARB, statin, diuretics, allopurinol, sodium, Hb, lymphocytes, UA, cholesterol, ethnicity, BMI, CCB, 26 different comorbidities	SHFM (improved)	NR/5044	5	0.83	0.82		0.8	Yes
Symptomatic with HFpEF										
Barlera, S. (2013) [47]	Cox model	Age, gender, SBP, NYHA class, BMI, ischemic HF etiology, smoking, COPD, DM, peripheral edema, HF hospitalizations, PVD, pacemaker, AS, hepatomegaly, THD, LVEF, QRS duration, AF, HR, eGFR, uricemia, Hb, triglycerides, fibrinogen	No	1969/6975	3.9			0.76 (0.75–0.77) *		Yes
Han, Y. (2023) [48]	Cox model	HF-PRS comprising 69 variants, age, COPD, SBP, NYHA class, BMI, NT-proBNP, sodium, BUN, creatinine, ACEi/ARB at discharge	No	212/1442	1	0.88 (0.86–0.9)				Yes
Kasahara, S. (2019) [50]	Cox model	Age, albumin, anemia, BMI, BNP/NT-proBNP, BUN	No, development of 3A3B	20/170	1	0.74 (0.63–0.86)				Yes

Table 2 (continued)

Study reference	Modelling method	Variables included in the final model	Use of previously developed risk scores	N. event/sample size	Median follow-up (years)	1-year *C*-statistic	2-year *C*-statistic	3-year *C*-statistic	5-year *C*-statistic	Validation (internal or external)
Mendez Fernandez, A.B. (2020) [51]	Cox model	Age, HF etiology, NYHA class, SBP, AF, left atrial diameter, eGFR, Hb, NT-proBNP, GDF-15 concentrations	No	98/311	1.3	0.82 (0.77–0.87)				Yes
Sun, Y. (2021) [52]	Cox model	Age, NYHA class, BMI, antihypertensive, paroxysmal or persistent AF, pulmonary arterial systolic pressure, E/e′ by Doppler echocardiography	H2FPEF	63/476	2.3		0.67 (0.6–0.73)			Yes
Angraal, S. (2020) [53]	Random Forest	Age, alkaline phosphatase level, BUN, BMI, KCCQ subscale scores	No	268/1088	3			0.72 (0.69–0.75)		Yes
Any										
Hammadah, M. (2014) [55]	Cox model	Age, gender, SBP, DM, HDL, LDL, BMI, smoking, creatinine and dialysis, ACEi, β-blockers, statins, nitrate, aspirin, MI, BNP, EF, ventricular rate, QRS duration, left bundle branch block, ICD placement, ceruloplasmin	No	261/890	5				0.7 (0.66–0.74)	Yes

Table 2 (continued)

Study reference	Modelling method	Variables included in the final model	Use of previously developed risk scores	N. event/ sample size	Median follow-up (years)	1-year C-statistic	2-year C-statistic	3-year C-statistic	5-year C-statistic	Validation (internal or external)
Honold, J. (2013) [56]	Cox model	Age, gender, cardiomyopathy etiology, HR, SBP, LVEF, ACEi, ARB, aldosterone blocker, β-blocker, statins, diuretic, allopurinol, sodium, cholesterol, Hb, lymphocytes, UA	SHFM	24/157	4			0.81 (0.71–0.90) *		Yes
Jackson, C.E. (2015) [57]	Cox model	Age, MI, NYHA class, BMI, smoking, BNP, bilirubin, hsCRP, Hb, lymphocyte, cFLC	No	290/628	3.2			0.73 (0.69–0.77)		Yes
Ky, B. (2012) [59]	Cox model	hsCRP, UA and MPO, BNP, sFlt-1, TnI, ST2, creatinine, SHFM score	SHFM (improved)	187/1513	2.5	0.81 (0.76–0.85)				Yes
Voors, A.A. (2017) [60]	Cox model	Age, Ischemic etiology, COPD, Peripheral edema, Elevated Jugular venous pressure, DBP, SBP, BUN, NT-proBNP, Hb, Ht, sodium, bilirubin, alkaline phosphatase, HDL	No	589/1738	2		0.73 (0.71–0.75)			Yes
Zafrir, B. (2012) [61]	Cox model	Age, gender, SBP, 6 MWD, β-blocker, hyperurcemia, hyponatremia, QTc interval	No	151/500	2		0.73 (0.72–0.74)			Yes

Table 2 (continued)

Study reference	Modelling method	Variables included in the final model	Use of previously developed risk scores	N. event/ sample size	Median follow-up (years)	1-year C-statistic	2-year C-statistic	3-year C-statistic	5-year C-statistic	Validation (internal or external)
Escobar, A. (2017) [62]	Logistic Regression	Age, SBP, NYHA class, heart valve disease, dementia, prior hospitalization, Hb, sodium, urea, length of hospital stay, physical dimension of Minnesota Living with HF questionnaire	No	295/1282	1	0.7 (0.66–0.73)				Yes
Williams, B.A. (2018) [63]^	Cox model	Age, gender, NYHA class, EF, ischemic etiology of HF, SBP, statin, allopurinol, diuretic, Hb, lymphocytes, UA, cholesterol, sodium	SHFM	14,380/26,851	4.7	0.66 (0.65–0.67)				Yes
Bowen, G.S. (2018)^ [64]	Logistic Regression	Age, BUN, sodium, BNP, potassium, EF, BP, metastasis, beats per minute, prior hospitalization, previous palliative care	No	11,980/30,111	2	0.68 (0.67–0.69)	0.67 (0.67–0.68)			Yes

Table 2 (continued)

Study reference	Modelling method	Variables included in the final model	Use of previously developed risk scores	N. event/sample size	Median follow-up (years)	1-year *C*-statistic	2-year *C*-statistic	3-year *C*-statistic	5-year *C*-statistic	Validation (internal or external)
Jing, L. (2020)^ [66]	Boosting	Age, gender, height, weight, smoking, HR, SBP, DBP, diuretic, anti-hypertensive, antidiabetic, Hb, eGFR, creatine kinase muscle/brain, lymphocytes, HDL, LDL, UA, sodium, potassium, NT-proBNP, troponin T, Hb, troponin I, creatinine, cholesterol, 90 cardiovascular-related ICD-10 codes, 44 nonredundant echocardiographic variables, 41 ECG measurements and patterns, 8 care gap variables	No	42/548	1	0.78 (0.71–0.85)				Yes
Sartipy, U. (2014)^ [67]	Cox model	Age, gender, EF, NYHA class, creatinine, diabetes, SBP, BMI, HF duration, current smoker, COPD, β-blocker, ACEi, ARB	MAGGIC heart failure risk score	21,936/51,043	2.6			0.74 (0.74–0.75)		Yes
Yap, J. (2019)^ [68]	Cox model	Age, prior MI, AF, prior stroke, PVD, LVEF, QRS duration, creatinine, SBP, sodium	No, development of Singapore Heart Failure Risk Score	262/729	2	0.67 (0.62–0.72)	0.68 (0.64–0.72)			Yes

Table 2 (continued)

Study reference	Modelling method	Variables included in the final model	Use of previously developed risk scores	N. event/ sample size	Median follow-up (years)	1-year *C*-statistic	2-year *C*-statistic	3-year *C*-statistic	5-year *C*-statistic	Validation (internal or external)
Quiros-Lopez, R. (2019)^ [69]	Logistic Regression	Age, SBP, NYHA class, valvular HF cause, dementia, prior HF hospitalization, Hb, sodium, urea, length of stay	CACE-HF	767/3337	1	0.67 (0.65–0.69)				Yes
Wang, L. (2012) [70]^	Cox model	Age, metastatic cancer, dementia, albumin, urea, HR, respiration rate	No	14,124/198,640	1	0.76 (0.75–0.76)				Yes
Wang, Z. (2018)^ [71]	Dynamic Radius Means	Age, gender, HR, diagnoses, medications, laboratory tests	No	1182/4353	1	0.85 (0.84–0.86)				Yes

HFpEF heart failure with preserved ejection fraction, *HFrEF* heart failure with reduced ejection fraction, *HFmrEF* heart failure with mid-range ejection fraction

[*] 4-year *C*-statistic

^studies using EHR data

Table 3 Most consistent and strongest predictors of risk of death across studies

Study reference	Age	RF	BP	CAD/IHD	Sodium	NYHA class	EF	Gender	Weight/BMI	BNP
Advanced HF										
Oh, C. (2012) [28]		x		x	x					
Regoli, F. (2013) [29]				x				x	x	
Richter, B. (2013) [30]	x									x
Scrutinio, D. (2014) [31]		x	x		x		x			x
Smith, T. (2012) [32]	x	x	x		x	x	x	x	x	
Ketchum, E.S. (2012) [33]	x	x	x	x	x	x	x	x		
O'Connor, C.M. (2012) [34]		x	x				x	x	x	
Simpson, J. (2020) [35]	x		x	x		x	x	x	x	
Upshaw, J.N. (2016) [36]	x	x	x	x	x	x	x	x	x	x
Zhang, J. (2013) [37]	x		x	x	x					x
Nymo, S.H. (2014) [38]										x
Li, H. (2020) [39]	x	x		x		x				x
Symptomatic or not with HFrEF + HFmrEF										
Clemens, M. (2012) [40]	x		x	x		x	x	x	x	
Vishram-Nielsen, J.K.K. (2020) [41]	x	x	x	x	x	x	x	x	x	
Laszczynska, O. (2016) [42]	x	x	x	x	x	x	x	x	x	
Bayes-Genis, A. (2014) [43]	x	x		x	x	x	x	x		x
Fontanive, P. (2013) [44]	x	x				x	x			x
Siriopol, D. (2021) [45]		x	x		x	x				
Panahiazar, M. (2016) [46]	x	x	x	x	x		x	x	x	
Symptomatic with HFpEF										
Barlera, S. (2013) [47]	x	x	x	x		x	x	x	x	
Han, Y. (2023) [48]	x	x	x		x	x			x	x
Kasahara, S. (2019) [50]	x	x							x	x
Mendez Fernandez, A.B. (2020) [51]	x	x	x	x		x				x
Sun, Y. (2021) [52]	x		x	x		x			x	
Angraal, S. (2020) [53]	x	x							x	
Any										
Hammadah, M. (2014) [55]	x	x	x	x			x	x	x	x
Honold, J. (2013) [56]	x	x	x	x	x		x	x		
Jackson, C.E. (2015) [57]	x			x		x			x	x
Ky, B. (2012) [59]		x								x
Voors, A.A. (2017) [60]	x	x	x	x	x					x
Zafrir, B. (2012) [61]	x	x	x		x			x		
Escobar, A. (2017) [62]	x	x	x	x	x	x				
Williams, B.A. (2018) [63]	x	x	x	x	x	x	x	x		
Bowen, G.S. (2018) [64]	x	x	x		x		x			
Jing, L. (2020) [66]	x	x	x	x	x			x	x	x
Sartipy, U. (2014) [67]	x	x	x			x	x	x	x	
Yap, J. (2019) [68]	x	x	x	x	x		x			
Quiros-Lopez, R. (2019) [69]	x	x	x	x	x	x				
Wang, L. (2012) [70]	x	x								
Wang, Z. (2018) [71]	x							x		
Total occurences as final variable	33	31	27	24	21	20	19	19	18	16
%	83%	78%	68%	60%	53%	50%	48%	48%	45%	40%
Advanced HF	62%	46%	62%	54%	38%	46%	54%	54%	46%	46%
Symptomatic or not with HFrEF + HFmrEF	83%	100%	67%	67%	83%	83%	83%	67%	50%	33%
Symptomatic with HFpEF	100%	83%	67%	50%	17%	67%	17%	17%	83%	50%
Any	93%	87%	73%	60%	60%	33%	40%	47%	27%	33%

Table 3 (continued)

BP blood pressure, *RF* renal function, *EF* ejection fraction, *BMI* body mass index, *CAD* coronary artery disease, *IHD* ischemic heart disease, *BNP* B-type natriuretic peptide, *NYHA* New York Heart Association, *HFpEF* heart failure with preserved ejection fraction, *HFrEF* heart failure with reduced ejection fraction, *HFmrEF* heart failure with mid-range ejection fraction

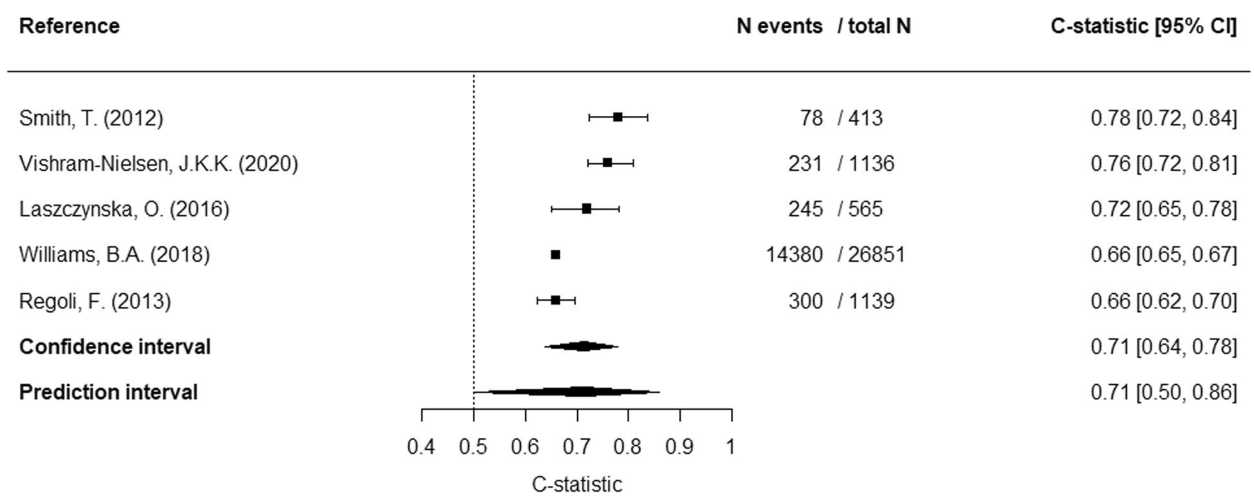

Fig. 3 Pooled *C*-statistic of the meta-analysis of the SHFMs for the prediction of 1-year risk of death

In detail, all the models underwent internal or external validation, and only 2 studies reported discrimination at 1 year (*C*-statistics 0.86, 95% CI: 0.82–0.89, in Bulluck et al. [65] and 0.82, 95% CI: 0.81–0.82, in Wang L. et al. [70]).

Interestingly, the study of Nadruz et al. [54] was the only one stratifying patients with respect to LVEF, and its predictive model for patients with an LVEF of 40% to 49% is the only one we found that focused only on patients with mildly reduced EF. In this study, HF-related hospitalizations were 31.7, 11.8, and 13.8% in patients with an LVEF < 40%, 40% to 49%, and ≥ 50%, respectively, resulting in 2-year discriminations of 0.70 (95% CI: 0.66–0.74), 0.74 (95% CI: 0.61–0.87), and 0.81 (95% CI: 0.72–0.9), respectively. This study was based on clinical data.

With respect to the modelling method, the study of Angraal et al. [53] was the only one that used a ML method (i.e., the random forest) for hospitalization and showed a discrimination in terms of *C*-statistics of 0.76 (95% CI: 0.71–0.81), with a Brier score of 0.19, for all-cause hospitalization at 3 years of follow-up.

Composite end-points

The remaining 9 models, derived from 8 studies, addressed composite end-points primarily combining death and hospitalization (Table 5). Most models focused on all-cause mortality (6 out of 9) rather than restricting the end-point to cardiovascular death (3 out of 9). With respect to hospitalization, 5 models included only HF-related hospitalizations, 3 considered all-cause hospitalizations, and one model by Nymo et al. [38] incorporated a broader composite end-point including cardiovascular mortality, nonfatal myocardial infarction (MI), and nonfatal stroke.

The percentage of composite events of death and hospitalization in the related 8 models ranged from 37.4% in the study of Nymo et al. [38], with a median follow-up of 2.7 years, to 67% in the study of O'Connor et al. [34], with a similar follow-up of 2.5 years. Only one study was based on mixed sources of data [70], resulting in a 1-year discrimination of 0.77 (95% CI: 0.77–0.77). At the same follow-up time (1 year), models based on clinical data reported higher discrimination: in particular, the study of Zhang et al. [37], which used a ML approach (i.e., the decision tree), reported a *C*-statistic of 0.80 (95% CI: 0.75–0.85), while the study of Kadowaki et al. [58], which employed a Cox model, found an apparent (not validated) *C*-statistics of 0.82 (95% CI: 0.75–0.90). At the following time points, only two studies were validated: O'Connor et al. [34] and Voors et al. [60] that reported 2-year discriminations of 0.64 (95% CI: 0.62–0.67) and 0.68 (95% CI: 0.66–0.7), respectively.

Discussion

Summary and interpretation of findings

This systematic review and meta-analysis provides a comprehensive overview of the prognostic models developed in patients with HF, including both HFpEF and

Table 4 Discrimination performance and main characteristics of models on hospitalization

Study reference	Modelling method	Variables included in the final model	*N*. event/sample size	Median follow-up (years)	1-year *C*-statistic	2-year *C*-statistic	3-year *C*-statistic	Validation (internal or external)	Note
Advanced HF									
Upshaw, J.N. (2016) [36]	Cox model	Age, gender, NYHA class, LVEF, creatinine, sodium, SBP, weight, DM, IHD, AF, prior stroke	NR/2521	3.8			0.62 *	Yes	*N*. of events not reported
Nadruz, W. (2017) [54]	Cox model	Age, gender, LVEF, CKD, SBP, HR, CAD, peakvo2, VE/VCO2 slope	200/630	2		0.7 (0.66–0.74)		Yes	LVEF < 40%
Symptomatic or not with HFrEF + HFmrEF									
Nadruz, W. (2017) [54]	Cox model	Age, gender, LVEF, CKD, SBP, HR, CAD, peakvo2	17/144	2		0.74 (0.61–0.87)		Yes	LVEF 40% to 49%
Symptomatic with HFpEF									
Sun, Y. (2021) [52]	Cox model	Age, NYHA class, BMI, antihypertensive, AF, pulmonary arterial systolic pressure, *E/e´* by Doppler echocardiography	311/476	2.3		0.59 (0.54–0.65)		Yes	
Angraal, S. (2020) [53]	Random Forest	Hb, BUN, time since previous HF hospitalization, KCCQ scores, eGFR, glucose	343/1088	3			0.76 (0.71–0.81)	Yes	
Nadruz, W. (2017) [54]	Cox model	Age, gender, LVEF, CKD, SBP, HR, CAD, peakvo2, VE/VCO2 slope	27/195	2		0.81 (0.72–0.9)		Yes	LVEF ≥ 50%
Any									
Voors, A.A. (2017) [60]	Cox model	Age, HF hospitalization, DM, COPD, peripheral edema, elevated jugular venous pressure, SBP, eGFR, NT-proBNP, HDL	610/1738	2		0.63 (0.6–0.66)		Yes	

Table 4 (continued)

Study reference	Modelling method	Variables included in the final model	*N*. event/sample size	Median follow-up (years)	1-year *C*-statistic	2-year *C*-statistic	3-year *C*-statistic	Validation (internal or external)	Note
Bulluck, H. (2019) [65]^	Logistic Regression	Age, IHD, DM, Killip class, creatinine, Hb, troponin, S2B time, LVEF	123/2280	1	0.86 (0.82–0.89)			Yes	
Wang, L. (2012) [70]^	Cox model	Prior hospitalizations, emergency visits, outpatient visits, high priority for VHA health care	46,919/198,640	1	0.82 (0.81–0.82)			Yes	

All models included HF-attributable hospitalization except Wang, L. (2012) [70] and Angraal, S. (2020) [53] that included all cause hospitalizations

HFpEF, heart failure with preserved ejection fraction; *HFrEF*, heart failure with reduced ejection fraction; *HFmrEF*, heart failure with mid-range ejection fraction

[*] 4-year C-statistic

^*studies using EHR data*

Table 5 Discrimination performance and main characteristics of the models on composite outcomes

Study reference	Modelling method	Variables included in the final model	*N*. event/ sample size	Median follow-up (years)	1-year *C*-statistic	2-year *C*-statistic	3-year *C*-statistic	5-year *C*-statistic	Validation (internal or external)	Note
Advanced HF										
O'Connor, C.M. (2012) [34]	Cox model	Gender, BUN, KCCQ symptom stability, KCCQ total symptom score, weber class, LVEF, exercise duration on CPX test, ventricular conduction prior to CPX, mitral regurgitation, race, treatment	1555/2331	2.5		0.64 (0.62–0.67)			Yes	All-cause death or all-cause hospitalization
Zhang, J. (2013) [37]	Decision Tree	NT-proBNP, sinus rhythm, MI, creatinine, urea, potassium	115/284	1	0.8 (0.75–0.85)				Yes	All-cause death or HF hospitalization
Nymo, S.H. (2014) [38]	Cox model	NT-proBNP, IL-8	320/1464	2.7			0.68 (0.65–0.71)		No	CV death, non-fatal MI, or non-fatal stroke
Nymo, S.H. (2014) [38]	Cox model	NT-proBNP, IL-8	547/1464	2.7			0.69 (0.67–0.72)		No	CV death or HF hospitalization
Symptomatic or not with HFrEF + HFmrEF										
Bayes-Genis, A. (2014) [43]	Cox model	Age, gender, LVEF, eGFR, NYHA class, DM, ischemic etiology, Hb, sodium, β-blocker, ACEi or ARB, NT-proBNP, ST2	NR/876	4.2				0.74 (0.72–0.77)	No	All-cause death or HF hospitalization
Symptomatic with HFpEF										
Kanagala, P. (2020) [49]	Cox model	Age, prior HF hospitalization, DBP, lung disease, NYHA, 6MWTD, Hb, creatinine, BNP, LAEF	67/140	3.9			0.81 (0.73–0.88) *		No	All-cause death or HF hospitalization

Table 5 (continued)

Study reference	Modelling method	Variables included in the final model	*N*. event/ sample size	Median follow-up (years)	1-year *C*-statistic	2-year *C*-statistic	3-year *C*-statistic	5-year *C*-statistic	Validation (internal or external)	Note
Any										
Kadowaki, S. (2016) [58]	Cox model	Age, BNP, eGFR, low BDNF	51/134	1.2	0.82 (0.75–0.9)				No	Cardiac death or HF hospitalization
Voors, A.A. (2017) [60]	Cox model	Age, prior HF hospitalization, smoking, COPD, NYHA class, peripheral Elevated Jugular venous pressure, SBP, eGFR, BUN, NT-proBNP, Hb, sodium, bilirubin, alkaline phosphatase, HDL	894/1738	2		0.68 (0.66–0.7)			Yes	All-cause death or all-cause hospitalization
Wang, L. (2012) [70]^	Cox model	Not reported	NR/198,640	1	0.77 (0.77–0.77)				Yes	All-cause death or all-cause hospitalization

HFpEF heart failure with preserved ejection fraction, *HFrEF* heart failure with reduced ejection fraction, *HFmrEF* heart failure with mid-range ejection fraction

* 4-year C-statistic

^*studies using EHR data*

HFrEF populations, and a range of outcomes including mortality, hospitalization, and composite endpoints. Overall, a wide range of models has been proposed, but important methodological limitations and suboptimal reporting were frequently observed. Thus, the real-world clinical applicability remains limited.

The heterogeneity of the HF patient population, particularly in terms of disease severity and symptom profile, was rarely explicitly addressed in prognostic studies, making the classification of studies by HF stage challenging. Most studies included patients across multiple stages, with only a minority focusing on more homogeneous cohorts, primarily patients in stage C of the Universal Definition of HF [1] or patients with reduced ejection fraction. Consequently, evidence supporting prognostic modelling in patients with preserved ejection fraction remains limited. This work also offers a systematic evaluation of methodological quality and risk of bias using PROBAST, identifying recurrent weakness across different domains. Calibration emerged as a key unmet need, as it was rarely reported. Only the Seattle Heart Failure Model was supported by a sufficient number of studies reporting discrimination at 1 year, allowing a meta-analysis to be performed. The pooled *C*-statistic of 0.71 (95% confidence interval: 0.64, 0.78; 95% prediction interval: 0.50, 0.86) not only indicates moderate discriminative ability but also highlights substantial variability across setting and populations.

Most models exhibited a moderate to high risk of bias, largely driven by methodological limitations in the analysis domain [24], while predictor and outcome assessment were generally well addressed in the included studies. The frequent use of univariable predictor selection and inadequate handling of missing data represent well-recognized sources of bias in prediction modelling studies and may compromise the predictive performance and generalizability of the models [72]. Notably, among the 58 models, only 7 reported calibration metrics, confirming calibration as the Achilles heel of predictive modelling [73]. Yet, calibration is crucial in risk prediction [20, 74, 75] as it ensures that predicted probabilities accurately reflect observed outcome rates and real-world risks. A model may show good discrimination while still providing inaccurate risk estimates, thereby limiting its reliability for clinical decision-making. Moreover, the absence of calibration information restricts the communication of individualized risk estimates to patients, limiting shared decision-making and appropriate patient counseling. Therefore, predicted risks should be interpreted with caution, particularly when models are used to support individualized decision-making or risk communication.

A key finding of this review is that the majority of studies did not use previously derived scores but instead developed new, study-specific models. Specifically, among the 58 identified prognostic models, 21 evaluated or validated an existing risk score, whereas 37 developed newly derived models. Among previously established scores, the SHFM was the most frequently evaluated. Most newly derived models relied on a largely overlapping set of core clinical and laboratory predictors. The predictors most frequently included were age, renal function, blood pressure, presence of CAD, serum sodium, NYHA class, ejection fraction, sex, weight or BMI, and B-type natriuretic peptide. Most of these predictors are also components of the SHFM. When the SHFM (incorporating demographic data, laboratory values, and information on device therapy and medication) was applied, the estimated discrimination at 1 year was 0.71 (95% confidence interval: 0.64, 0.78; 95% prediction interval: 0.50, 0.86), indicating acceptable discrimination with moderate heterogeneity, partially attributable to differences in patient populations. Notably, most SHFM validations were performed in patients with reduced ejection fraction, while its applicability to preserved ejection fraction populations remains mostly unproven. Moreover, the SHFM, published in 2006, was derived mainly from cohorts enrolled before the widespread adoption of implantable cardioverter-defibrillators (ICDs) and cardiac resynchronization therapy (CRT). Consequently, it reflects pre-device-era mortality patterns, limiting its temporal validity and generalizability to contemporary HF populations.

Among the 58 models included in this systematic review, only three explicitly incorporated device therapy (ICD or CRT) among their final predictors—namely, those by Hammadah et al. [55], Ketchum et al. [33], and Regoli et al. [29], the latter including a CRT-P device. Most models were developed between 2012 and 2020, when the adoption of ICD/CRT and contemporary pharmacotherapy (e.g., ARNI, SGLT2 inhibitors) was still evolving. This is also a limit of our literature search that was conducted through to 2023, but we expect future model updates to explicitly incorporate contemporary therapies and evaluate their incremental prognostic contribution [76, 77].

Our findings also highlight a narrow conceptualization of prognosis in existing models, which focused predominantly on mortality (40 out of 58 models). Other clinically relevant outcomes, such as hospitalization, were less frequently considered. A subset of models (9 models from 8 studies) used composite endpoints combining death and hospitalization, and in one case, stroke. Composite endpoints can increase event rates and improve model discrimination, but they also introduce heterogeneity because components differ in clinical relevance. Overall, models using composite endpoints showed reasonable short-term discrimination, although longer-term validation was limited. Composite endpoints are acceptable in this setting [78], but multidimensional prognostic approaches, such as ordinal longitudinal models [79], may better integrate survival with quality-of-life and functional trajectories. This approach is particularly relevant in HF populations, where frailty is common and strongly influences patient-centered outcomes and clinical decision-making [80, 81].

Statistical models remain more common than machine learning (ML) approaches (88% versus 12%). Although direct comparison was not possible, available evidence suggests that ML models (e.g., random forest, decision trees or boosting) did not consistently outperform traditional statistical approaches (e.g., Cox models and logistic regression) in predicting HF outcomes. Overall, ML models achieved similar performance in predicting all-cause mortality or HF-related hospitalization, depending on the type and richness of predictors. At present, ML methods have not demonstrated sufficient added value to substantially change prognostic modelling in HF, as also reported by the systematic review of Sun et al. [14], although their potential is of interest. A recent review focusing on ML methods for predicting HF survival highlighted superior performance of ML algorithms compared with traditional statistical approaches [82]. However, these comparisons are often flawed because statistical regression models are applied without adequately accounting for the complexity of the data. Additionally, ML models frequently suffer from issues such as data leakage, leading to overly optimistic performance estimates [83]. Furthermore, ML-based models continue to face challenges related to their limited interpretability, restricting their clinical applicability.

Finally, model usability and implementation aspects were rarely addressed. Many models require numerous or non-routinely available predictors, limiting ease of use, while few studies discussed integration into clinical workflows or electronic health records. These practical limitations further restrict the translation of existing prognostic models into real-world clinical practice.

Implications for future prognostic model development

The limitations of the studies included in this systematic review warrant careful consideration. Substantial clinical variability among patients and methodological heterogeneity across studies limited the comparability of predictive performance and complicated meta-analytic synthesis. The only meta-analysis performed was based on a limited number of studies and was subject to potential risk of bias in the original studies. Accordingly, summary estimates of predictive performance should be interpreted with caution. These limitations underscore the need for transparent and standardized methodological practices in the design and reporting of prognostic research in this field. This is particularly relevant for ML techniques, as their application in medicine is still relatively early and pioneering studies may be methodologically immature. Further work is required to clarify how these approaches can be robustly developed, validated, and implemented in clinical settings [14].

Future prognostic models should adhere to transparent reporting standards such as TRIPOD and TRIPOD-AI [74, 75], ensuring full transparency of model specifications and performance metrics. Adequate study size is essential to develop and validate reliable prognostic models [84] and to minimize overfitting. Rigorous methods for model development should be employed; for example, strong internal validation using resampling techniques that repeat all modelling steps may prevent expensive and time-consuming validation of poorly performing models [72]. External validation should routinely assess both discrimination and calibration in independent cohorts, a practice that was rarely observed in the included studies. Validation across diverse geographic and temporal settings should be a prerequisite for clinical adoption [20, 74, 85]. Models should also be periodically updated to reflect evolving therapeutic scenarios [86], including the impact of contemporary pharmacological and device-based therapies on prognosis [84, 87]. Future models may further benefit from including psychosocial and behavioral predictors, frailty indicators, and patient-reported outcomes [78, 88–91].

It is also important to ensure adequate representation of under-studied groups, including older adults, sex-based subgroups, patients with preserved ejection fraction, and multimorbid patients. Explicit inclusion of these populations in both derivation and validation cohorts is essential to improve generalizability [92].

Emerging data sources such as wearable sensors and home telemonitoring offer the potential to improve prognostic models by capturing longitudinal trajectories of physiological parameters, functional status, and patient-reported outcomes within dynamic prediction frameworks [93–95]. However, their integration requires standardization, prospective validation, and robust analytic methods to manage high-frequency and noisy data [96, 97]. While these approaches introduce challenges related to data quality, privacy, interoperability, and model complexity, they also represent a promising path for improving predictive performance [98]. Artificial intelligence and ML methods may facilitate the identification of complex, non-linear relationships and dynamic patterns [99, 100], but they should be applied within transparent, reproducible frameworks and subjected to the same rigorous validation standards as traditional models [101] acknowledging their typically larger sample size requirements [102, 103].

Finally, to translate prognostic modelling into tangible clinical benefit, collaborative frameworks involving clinicians, statisticians, data scientists, and policymakers are needed. Prospective studies evaluating the clinical impact of model-guided decision-making on patient outcomes should represent the next frontier of research in heart failure prediction [20, 74].

Conclusions

In summary, this systematic review and meta-analysis shows that, although numerous prognostic models for HF have been developed, most exhibit substantial methodological limitations that restrict their clinical applicability. At present, the real-world usefulness of most models for individual risk stratification and clinical decision-making remains limited. Discrimination was generally moderate, as illustrated by the pooled *C*-statistic of 0.71 for the SHFM at 1 year, but calibration was poorly reported in the majority of studies, limiting the assessment of absolute risk predictions and generalizability.

Considerable heterogeneity was observed across studies in patient populations, outcome definitions, follow-up durations, and modeling approaches, including both traditional statistical models and ML methods.

Among the evaluated models, the SHFM remains the tool with the strongest evidence and was the only model suitable for meta-analysis; however, its clinical utility is constrained by incomplete calibration reporting and limited applicability to contemporary therapies and diverse patient phenotypes. Most other models also lacked external validation, calibration, representation of diverse HF phenotypes (including preserved ejection fraction), and alignment with contemporary therapeutic contexts.

Future efforts should prioritize transparent reporting, robust external validation across diverse cohorts, inclusiveness across patient phenotypes, incorporation of current therapeutic strategies, and the development of dynamic, multidimensional, patient-centered models that integrate clinical, functional, and patient-reported outcomes. Ultimately, improving methodological rigor and aligning model development with evolving therapies and data innovations will enhance the potential for prognostic models to guide personalized management and decision-making in heart failure.

Abbreviations

AUC	Area under the ROC Curve
BMI	Body mass index
CAD	Coronary artery disease
CCB	Calcium channel blocker
EF	Ejection fraction
EHR	Electronic health records
HF	Heart failure
HFmrEF	Heart failure with mildly reduced ejection fraction
HFpEF	Heart failure with preserved ejection fraction
HFrEF	Heart failure with reduced ejection fraction
LVEF	Left ventricular ejection fraction
MACE	Major adverse cardiovascular event
MI	Myocardial infarction
ML	Machine learning
NYHA	New York Heart Association
PROBAST	Prediction Model Risk of Bias Assessment Tool
RCT	Randomized controlled trial
RoB	Risk of bias
ROC	Receiver operating characteristic
SHFM	Seattle Heart Failure Model
WHF	Worsening heart failure

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s13643-026-03100-5.

Additional file 1: PRISMA Checklist.

Additional file 2: Search string.

Additional file 3: PROBAST risk of bias and applicability for each domain in each model.

Acknowledgements

The authors would like to thank the librarian of the University of Milano-Bicocca, Laura Colombo, for assisting in developing the literature search strategy. Working group: PROPHET-I (PRediction Of chronic disease Progression through Health Electronic and Territorial data in Italian populations). Study group composition: Anita Andreano (1), Lucia Bisceglia (2), Martino Bussa (3), Endeshaw Assefa Derso (3), Caterina Fanizza (2), Dariush Khaleghi Hashemian (1), Vito Lepore (2), Pietro Magnoni (1), Cristina Mazzali (1), Alberto Milanese (1), Alessandro Musa (2), Giuseppe Occhino (2), Matteo Petrosino (2), Paola Rebora (3), Antonio Giampiero Russo (1), Deborah Testa (1), Sara Tunesi (1), Maria Grazia Valsecchi (3), Adele Zanfino (1).
(1) Agenzia di Tutela della Salute della Città Metropolitana di Milano – SC Unità di Epidemiologia; (2) Agenzia Regionale Strategica per la Salute ed il Sociale (AReSS) Puglia – Area Epidemiologia e Care Intelligence; (3) Università Milano Bicocca – Dipartimento di Medicina e Chirurgia e Centro Interdipartimentale Bicocca Bioinformatics Biostatistics and Bioimaging Centre (B4).

Authors' contributions

PR and AA contributed to study design. PR, AA, and GO contributed to the protocol's planning and creation. GO and AM performed the bibliographic searches, screened titles and abstracts, assessed full papers, and extracted the data. GO, PR, and MB performed data analysis, drafted the manuscript. GO, MB, PR, LB, and AGR contributed to data interpretation. All authors revised and approved the final version of the manuscript. PR, LB, and AGR acted as the guarantor of the review.

Funding

Funded by the European Union–Next Generation EU–NRRP M6C2–Investment 2.1 Enhancement and strengthening of biomedical research in the NHS project code PNRR-MAD-2022-12376033, title "Evidence-based models for high impact chronic disease prevention and risk of progression management in outpatient community services and community hospitals: toward eHealth integrating stratification on individual history with predictive models of disease progression, using machine learning and artificial intelligence on administrative and clinical databases," PI Antonio Giampiero Russo.

Data availability

The datasets underlying this article will be shared upon reasonable request to the corresponding author.

Declarations

Ethics approval and consent to participate

The study concerns literature-based studies. Therefore, ethical approval and informed consent were not needed.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Author details
[1]Bicocca Bioinformatics Biostatistics and Bioimaging (B4) Centre, School of Medicine and Surgery, University of Milano-Bicocca, Monza, Italy. [2]Strategic Regional Agency for Health and Social Care of Apulia (AReSS Puglia), Bari, Italy. [3]Biostatistics and Clinical Epidemiology, Fondazione IRCCS San Gerardo Dei Tintori, Monza, Italy. [4]Epidemiology Unit, Agency for Health Protection of Milan (ATS Milano), Milan, Italy.

Received: 27 January 2025 Accepted: 3 February 2026
Published online: 19 February 2026

References

1. Bozkurt B, Coats AJ, Tsutsui H, et al. Universal definition and classification of heart failure. J Card Fail. 2021;27(4):387–413. https://doi.org/10.1016/j.cardfail.2021.01.022.
2. McDonagh TA, Metra M, et al. 2021 ESC guidelines for the diagnosis and treatment of acute and chronic heart failure: developed by the Task Force for the diagnosis and treatment of acute and chronic heart failure of the European Society of Cardiology (ESC). With the special contribution of the Heart Failure Association (HFA) of the ESC. Eur J Heart Fail. 2022;24(1):4–131. https://doi.org/10.1002/ejhf.2333.
3. Bytyçi I, Bajraktari G. Mortality in heart failure patients. Anadolu Kardiyol Derg-Anatol J Cardiol. 2015;15(1):63–8. https://doi.org/10.5152/akd.2014.5731.
4. Horne BD, Roberts CA, Rasmusson KD, et al. Risk score–guided multidisciplinary team-based care for Heart Failure Inpatients is associated with lower 30-day readmission and lower 30-day mortality. Am Heart J. 2020;219:78–88. https://doi.org/10.1016/j.ahj.2019.09.004.
5. Seferović P, Polovina M, Milinković I, Krljanac G, Ašanin M. Risk stratification models for predicting mortality in heart failure: a favourite or an outsider? Eur J Prev Cardiol. 2024;31(2):272–3. https://doi.org/10.1093/eurjpc/zwac173.
6. Ammar LA, Massoud GP, Chidiac C, Booz GW, Altara R, Zouein FA. BNP and NT-proBNP as prognostic biomarkers for the prediction of adverse outcomes in HFpEF patients: a systematic review and meta-analysis. Heart Fail Rev. 2024;30(1):45–54. https://doi.org/10.1007/s10741-024-10442-6.
7. Levy WC, Mozaffarian D, Linker DT, et al. The Seattle Heart Failure Model: prediction of survival in heart failure. Circulation. 2006;113(11):1424–33. https://doi.org/10.1161/CIRCULATIONAHA.105.584102.
8. Pocock SJ, Ariti CA, McMurray JJV, et al. Predicting survival in heart failure: a risk score based on 39 372 patients from 30 studies. Eur Heart J. 2013;34(19):1404–13. https://doi.org/10.1093/eurheartj/ehs337.
9. Codina P, Lupón J, Borrellas A, et al. Head-to-head comparison of contemporary heart failure risk scores. Eur J Heart Fail. 2021;23(12):2035–44. https://doi.org/10.1002/ejhf.2352.
10. Kondo T, Jering KS, Jhund PS, et al. Predicting stroke in heart failure and preserved ejection fraction without atrial fibrillation. Circ Heart Fail. 2023;16(7):e010377. https://doi.org/10.1161/CIRCHEARTFAILURE.122.010377.
11. Rahimi K, Bennett D, Conrad N, et al. Risk prediction in patients with heart failure. JACC: Heart Fail. 2014;2(5):440–6. https://doi.org/10.1016/j.jchf.2014.04.008.
12. Amini-Rarani M, Karimi SE, SoleimanvandiAzar N, Nosratabadi M. Risk and protective factors related to having a child with congenital heart diseases (CHD): a case-control study in Iran. Front Pediatr. 2023;11:1170743. https://doi.org/10.3389/fped.2023.1170743.
13. Mpanya D, Celik T, Klug E, Ntsinjana H. Predicting mortality and hospitalization in heart failure using machine learning: a systematic literature review. IJC Heart Vasc. 2021;34:100773. https://doi.org/10.1016/j.ijcha.2021.100773.
14. Sun Z, Dong W, Shi H, Ma H, Cheng L, Huang Z. Comparing machine learning models and statistical models for predicting heart failure events: a systematic review and meta-analysis. Front Cardiovasc Med. 2022;9:812276. https://doi.org/10.3389/fcvm.2022.812276.
15. Lumish HS, Liang LW, Hasegawa K, et al. Prediction of worsening heart failure in hypertrophic cardiomyopathy using plasma proteomics. Heart. 2023. https://doi.org/10.1136/heartjnl-2023-322644.
16. Siddiqi TJ, Ahmed A, Greene SJ, et al. Performance of current risk stratification models for predicting mortality in patients with heart failure: a systematic review and meta-analysis. Eur J Prev Cardiol. 2022;29(15):2027–48. https://doi.org/10.1093/eurjpc/zwac148.
17. Di Tanna GL, Wirtz H, Burrows KL, Globe G. Evaluating risk prediction models for adults with heart failure: a systematic literature review. PLoS One. 2020;15(1):e0224135. https://doi.org/10.1371/journal.pone.0224135.
18. Xu Z, Hu Y, Shao X, et al. The efficacy of machine learning models for predicting the prognosis of heart failure: a systematic review and meta-analysis. Cardiology. 2024. https://doi.org/10.1159/000538639
19. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ. 2021:n71. https://doi.org/10.1136/bmj.n71
20. Debray TPA, Damen JAAG, Snell KIE, et al. A guide to systematic review and meta-analysis of prediction model performance. BMJ. 2017:i6460. https://doi.org/10.1136/bmj.i6460
21. Kohl C, McIntosh EJ, Unger S, et al. Online tools supporting the conduct and reporting of systematic reviews and systematic maps: a case study on CADIMA and review of existing tools. Environ Evid. 2018;7(1):8. https://doi.org/10.1186/s13750-018-0115-5.
22. Moons KGM, De Groot JAH, Bouwmeester W, et al. Critical appraisal and data extraction for systematic reviews of prediction modelling studies: the CHARMS checklist. PLoS Med. 2014;11(10):e1001744. https://doi.org/10.1371/journal.pmed.1001744.
23. Newcombe RG. Confidence intervals for an effect size measure based on the Mann–Whitney statistic. Part 2: asymptotic methods and evaluation. Stat Med. 2006;25(4):559–73. https://doi.org/10.1002/sim.2324.
24. Wolff RF, Moons KGM, Riley RD, et al. PROBAST: a tool to assess the risk of bias and applicability of prediction model studies. Ann Intern Med. 2019;170(1):51. https://doi.org/10.7326/M18-1376.
25. Damen JAA, Moons KGM, Van Smeden M, Hooft L. How to conduct a systematic review and meta-analysis of prognostic model studies. Clin Microbiol Infect. 2023;29(4):434–40. https://doi.org/10.1016/j.cmi.2022.07.019.
26. Higgins JPT, Thomas J, Chandler J, et al., eds. Cochrane handbook for systematic reviews of interventions. 1st ed. Wiley; 2019. https://doi.org/10.1002/9781119536604
27. Rücker G, Schwarzer G, Carpenter JR, Schumacher M. Undue reliance on I2 in assessing heterogeneity may mislead. BMC Med Res Methodol. 2008;8(1):79. https://doi.org/10.1186/1471-2288-8-79.
28. Oh C, Chang HJ, Sung JM, Kim JY, Yang W, Shim J, Kang SM, Ha J, Rim SJ, Chung N. Prognostic Estimation of Advanced Heart Failure With Low Left Ventricular Ejection Fraction and Wide QRS Interval. Korean Circ J. 2012;42(10):659–67. https://doi.org/10.4070/kcj.2012.42.10.659.
29. Regoli F, Scopigni F, Leyva F, et al. Validation of Seattle Heart Failure Model for mortality risk prediction in patients treated with cardiac resynchronization therapy. Eur J Heart Fail. 2013;15(2):211–20. https://doi.org/10.1093/eurjhf/hfs162.
30. Richter B, Koller L, Hohensinner PJ, Zorn G, Brekalo M, Berger R, Mörtl D, Maurer G, Pacher R, Huber K, Wojta J, Hülsmann M, Niessner A. A multi-biomarker risk score improves prediction of long-term mortality in patients with advanced heart failure. Int J Cardiol. 2013;168(2):1251–7. https://doi.org/10.1016/j.ijcard.2012.11.052.
31. Scrutinio D, Ammirati E, Guida P, Passantino A, Raimondo R, Guida V, Sarzi Braga S, Canova P, Mastropasqua F, Frigerio M, Lagioia R, Oliva F. The ADHF/NT-proBNP risk score to predict 1-year mortality in hospitalized patients with advanced decompensated heart failure. J Heart Lung Transplant. 2014;33(4):404–11. https://doi.org/10.1016/j.healun.2013.12.005.
32. Smith T. Performance of the Seattle Heart Failure Model in implantable defibrillator patients treated with cardiac resynchronization therapy. Am J Cardiol. 2012;110(3):398–402. https://doi.org/10.1016/j.amjcard.2012.03.038.
33. Ketchum ES, Jacobson AF, Caldwell JH, et al. Selective improvement in Seattle Heart Failure Model risk stratification using iodine-123

meta-iodobenzylguanidine imaging. J Nucl Cardiol. 2012;19(5):1007–16. https://doi.org/10.1007/s12350-012-9603-0.
34. O'Connor CM. Factors related to morbidity and mortality in patients with chronic heart failure with systolic dysfunction the HF-ACTION predictive risk score model. Circ Heart Fail. 2012;5(1):63–71.
35. Simpson J, Jhund PS, Lund LH, Padmanabhan S, Claggett BL, Shen L, Petrie MC, Abraham WT, Desai AS, Dickstein K, Køber L, Packer M, Rouleau JL, Mueller-Velten G, Solomon SD, Swedberg K, Zile MR, McMurray JJV. Prognostic Models Derived in PARADIGM-HF and Validated in ATMOSPHERE and the Swedish Heart Failure Registry to Predict Mortality and Morbidity in Chronic Heart Failure. JAMA Cardiol. 2020;5(4):432–41. https://doi.org/10.1001/jamacardio.2019.5850. Erratum in: JAMA Cardiol. 2020;5(4):488. https://doi.org/10.1001/jamacardio.2020.0451.
36. Upshaw JN, Konstam MA, Klaveren Dv, Noubary F, Huggins GS, Kent DM. Multistate Model to Predict Heart Failure Hospitalizations and All-Cause Mortality in Outpatients With Heart Failure With Reduced Ejection Fraction: Model Derivation and External Validation. Circ Heart Fail. 2016;9(8):e003146. https://doi.org/10.1161/CIRCHEARTFAILURE.116.003146.
37. Zhang J. Identifying patients at risk of death or hospitalisation due to worsening heart failure using decision tree analysis: evidence from the Trans-European Network-Home-Care Management System (TEN-HMS) study. Int J Cardiol. 2013;163(2):149–56.
38. Nymo SH, Hulthe J, Ueland T, et al. Inflammatory cytokines in chronic heart failure: interleukin-8 is associated with adverse outcome. results from CORONA. Eur J Heart Fail. 2014;16(1):68–75. https://doi.org/10.1093/eurjhf/hft125.
39. Li H, Cui Y, Tian J, Yang H, Zhang Q, Wang K, Han Q, Zhang Y. A novel age-biomarker-clinical history prognostic index for heart failure with reduced left ventricular ejection fraction. Open Med (Wars). 2020;15(1):644–53. https://doi.org/10.1515/med-2020-0209.
40. Clemens M, Szegedi Z, Kardos L, Nagy-Baló E, Sándorfi G, Edes I, Csanádi Z. The Seattle Heart Failure Model predicts survival in patients with cardiac resynchronization therapy: a validation study. J Card Fail. 2012;18(9):682–7. https://doi.org/10.1016/j.cardfail.2012.07.001.
41. Vishram-Nielsen JKK. Performance of prognostic risk scores in heart failure patients: do sex differences exist? Can J Cardiol. 2020;36(1):45–53. https://doi.org/10.1016/j.cjca.2019.08.021.
42. Laszczynska O. Validity of the Seattle Heart Failure Model for prognosis in a population at low coronary heart disease risk. J Cardiovasc Med. 2016;17(9):653–8. https://doi.org/10.2459/jcm.0000000000000048.
43. Bayes-Genis A, de Antonio M, Vila J, Peñafiel J, Galán A, Barallat J, Zamora E, Urrutia A, Lupón J. Head-to-head comparison of 2 myocardial fibrosis biomarkers for long-term heart failure risk stratification: ST2 versus galectin-3. J Am Coll Cardiol. 2014;63(2):158–66. https://doi.org/10.1016/j.jacc.2013.07.087.
44. Fontanive P, Miccoli M, Simioniuc A, Angelillis M, Di Bello V, Baggiani A, Bongiorni MG, Marzilli M, Dini FL. A multiparametric clinical and echocardiographic score to risk stratify patients with chronic systolic heart failure: derivation and testing. Echocardiography. 2013;30(10):1172–9. https://doi.org/10.1111/echo.12255.
45. Siriopol D, Popa R, Mihaila M, Rusu F, Sascau R, Statescu C, Cătălina Z, Vasiliu V, Bucur A, Neamtu A, Siriopol I, Cianga P, Kanbay M, Covic A. Application of survival classification and regression tree analysis for identification of subgroups of risk in patients with heart failure and reduced left ventricular ejection fraction. Int J Cardiovasc Imaging. 2021;37(6):1853–61. https://doi.org/10.1007/s10554-021-02159-6.
46. Panahiazar M, Taslimitehrani V, Pereira N, Pathak J. Using EHRs and Machine Learning for Heart Failure Survival Analysis. Stud Health Technol Inform. 2015;216:40–4.
47. Barlera S, Tavazzi L, Franzosi MG, Marchioli R, Raimondi E, Masson S, Urso R, Lucci D, Nicolosi GL, Maggioni AP, Tognoni G; GISSI-HF Investigators. Predictors of mortality in 6975 patients with chronic heart failure in the Gruppo Italiano per lo Studio della Streptochinasi nell'Infarto Miocardico-Heart Failure trial: proposal for a nomogram. Circ Heart Fail. 2013;6(1):31–9. https://doi.org/10.1161/CIRCHEARTFAILURE.112.967828.
48. Han Y, Lu J, Chen B, Li X, Dai H, Zhang L, Yan X, Liu J, Zhang H, Fu X, Yu Q, Ren J, Cui H, Gao Y, Li J. A novel polygenic risk score improves prognostic prediction of heart failure with preserved ejection fraction in the Chinese Han population. Eur J Prev Cardiol. 2023;30(13):1382–90. https://doi.org/10.1093/eurjpc/zwad209.
49. Kanagala P, Arnold JR, Cheng ASH, Singh A, Khan JN, Gulsin GS, Yang J, Zhao L, Gupta P, Squire IB, Ng LL, McCann GP. Left atrial ejection fraction and outcomes in heart failure with preserved ejection fraction. Int J Cardiovasc Imaging. 2020;36(1):101–10. https://doi.org/10.1007/s10554-019-01684-9.
50. Kasahara S, Sakata Y, Nochioka K, Tay WT, Claggett BL, Abe R, Oikawa T, Sato M, Aoyanagi H, Miura M, Shiroto T, Takahashi J, Sugimura K, Teng TK, Miyata S, Shimokawa H. The 3A3B score: The simple risk score for heart failure with preserved ejection fraction - A report from the CHART-2 Study. Int J Cardiol. 2019;284:42–9. https://doi.org/10.1016/j.ijcard.2018.10.076.
51. Mendez Fernandez AB, Ferrero-Gregori A, Garcia-Osuna A, Mirabet-Perez S, Pirla-Buxo MJ, Cinca-Cuscullola J, Ordonez-Llanos J, Roig Minguell E. Growth differentiation factor 15 as mortality predictor in heart failure patients with non-reduced ejection fraction. ESC Heart Fail. 2020;7(5):2223–9. https://doi.org/10.1002/ehf2.12621.
52. Sun Y, Wang N, Li X, Zhang Y, Yang J, Tse G, Liu Y. Predictive value of H2 FPEF score in patients with heart failure with preserved ejection fraction. ESC Heart Fail. 2021;8(2):1244–52. https://doi.org/10.1002/ehf2.13187.
53. Angraal S. Machine learning prediction of mortality and hospitalization in heart failure with preserved ejection fraction. JACC: Heart Failure. 2020;8(1):12–21.
54. Nadruz WJ, West E, Sengeløv M, et al. Prognostic value of cardiopulmonary exercise testing in heart failure with reduced, midrange, and preserved ejection fraction. J Am Heart Assoc. 2017. https://doi.org/10.1161/JAHA.117.006000.
55. Hammadah M, Fan Y, Wu Y, Hazen SL, Tang WHW. Prognostic value of elevated serum ceruloplasmin levels in patients with heart failure. J Card Fail. 2014. https://doi.org/10.1016/j.cardfail.2014.08.001.
56. Honold J, DeRosa S, Spyridopoulos I, Fischer-Rasokat U, Seeger FH, Leistner D, Lotz S, Levy WC, Zeiher AM, Assmus B. Comparison of the Seattle heart failure model and cardiopulmonary exercise capacity for prediction of death in patients with chronic ischemic heart failure and intracoronary progenitor cell application. Clin Cardiol. 2013;36(3):153–9. https://doi.org/10.1002/clc.22093.
57. Jackson CE, Haig C, Welsh P, Dalzell JR, Tsorlalis IK, McConnachie A, Preiss D, McInnes IB, Sattar N, Petrie MC, Gardner RS, McMurray JJ. Combined Free Light Chains Are Novel Predictors of Prognosis in Heart Failure. JACC Heart Fail. 2015;3(8):618–25. https://doi.org/10.1016/j.jchf.2015.03.014.
58. Kadowaki S, Shishido T, Honda Y, et al. Additive clinical value of serum brain-derived neurotrophic factor for prediction of chronic heart failure outcome. Heart Vessels. 2016;31(4):535–44. https://doi.org/10.1007/s00380-015-0628-6.
59. Ky B, French B, Levy WC, Sweitzer NK, Fang JC, Wu AH, Goldberg LR, Jessup M, Cappola TP. Multiple biomarkers for risk prediction in chronic heart failure. Circ Heart Fail. 2012;5(2):183–90. https://doi.org/10.1161/CIRCHEARTFAILURE.111.965020.
60. Voors AA. Development and validation of multivariable models to predict mortality and hospitalization in patients with heart failure. Eur J Heart Fail. 2017;19(5):627–34.
61. Zafrir B, Goren Y, Paz H, Wolff R, Salman N, Merhavi D, Lavi I, Lewis BS, Amir O. Risk score model for predicting mortality in advanced heart failure patients followed in a heart failure clinic. Congest Heart Fail. 2012;18(5):254–61. https://doi.org/10.1111/j.1751-7133.2012.00286.x.
62. Escobar A, García-Pérez L, Navarro G, Bilbao A, Quiros R; CACE-HF Score group. A one-year mortality clinical prediction rule for patients with heart failure. Eur J Intern Med. 2017;44:49–54. https://doi.org/10.1016/j.ejim.2017.06.013.
63. Williams BA, Agarwal S. Applying the Seattle Heart Failure Model in the office setting in the era of electronic medical records. Circ J. 2018;82(3):724.
64. Bowen GS, Diop MS, Jiang L, Wu WC, Rudolph JL. A Multivariable Prediction Model for Mortality in Individuals Admitted for Heart Failure. J Am Geriatr Soc. 2018;66(5):902–8. https://doi.org/10.1111/jgs.15319.
65. Bulluck H, Zheng H, Chan MY, et al. Independent Predictors of Cardiac Mortality and Hospitalization for Heart Failure in a Multi-Ethnic Asian ST-segment Elevation Myocardial Infarction Population Treated by Primary Percutaneous Coronary Intervention. Sci Rep. 2019;9:10072. https://doi.org/10.1038/s41598-019-46486-0.

66. Jing L, Ulloa Cerna AE, Good CW, Sauers NM, Schneider G, Hartzel DN, Leader JB, Kirchner HL, Hu Y, Riviello DM, Stough JV, Gazes S, Haggerty A, Raghunath S, Carry BJ, Haggerty CM, Fornwalt BK. A Machine Learning Approach to Management of Heart Failure Populations. JACC Heart Fail. 2020;8(7):578–87. https://doi.org/10.1016/j.jchf.2020.01.012.
67. Sartipy U, Dahlström U, Edner M, Lund LH. Predicting survival in heart failure: validation of the MAGGIC heart failure risk score in 51,043 patients from the Swedish heart failure registry. Eur J Heart Fail. 2014;16(2):173–9. https://doi.org/10.1111/ejhf.32.
68. Yap J, Chia SY, Lim FY, Allen JC, Teo L, Sim D, Go YY, Jaufeerally FR, Seow M, Kwok B, Liew R, Lam CS, Ching CK. The Singapore Heart Failure Risk Score: Prediction of Survival in Southeast Asian Patients. Ann Acad Med Singap. 2019;48(3):86–94.
69. Quirós-López R, Manzano-Espinosa L, Bilbao A, Montero Pérez-Barquero M, Cepeda-Rodrigo JM, Escobar-Martínez A. External validation of the CACE-HF risk score for mortality in patients with heart failure. Eur J Intern Med. 2019;66:35–40. https://doi.org/10.1016/j.ejim.2019.05.010.
70. Wang L. Predicting risk of hospitalization or death among patients with heart failure in the Veterans Health Administration. Am J Cardiol. 2012;110(9):1342–9.
71. Wang Z, Yao L, Li D, Ruan T, Liu M, Gao J. Mortality prediction system for heart failure with orthogonal relief and dynamic radius means. Int J Med Inform. 2018;115:10–7. https://doi.org/10.1016/j.ijmedinf.2018.04.003.
72. Harrell FE. Regression Modeling Strategies: With Applications to Linear Models, Logistic and Ordinal Regression, and Survival Analysis. Springer International Publishing; 2015. https://doi.org/10.1007/978-3-319-19425-7.
73. 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. https://doi.org/10.1186/s12916-019-1466-7.
74. Moons KGM, Altman DG, Reitsma JB, et al. Transparent reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD): explanation and elaboration. Ann Intern Med. 2015;162(1):W1–73. https://doi.org/10.7326/M14-0698.
75. Collins GS, Dhiman P, Andaur Navarro CL, et al. Protocol for development of a reporting guideline (TRIPOD-AI) and risk of bias tool (PROBAST-AI) for diagnostic and prognostic prediction model studies based on artificial intelligence. BMJ Open. 2021;11(7):e048008. https://doi.org/10.1136/bmjopen-2020-048008.
76. McDonagh TA, Metra M, Adamo M, et al. 2023 focused update of the 2021 ESC guidelines for the diagnosis and treatment of acute and chronic heart failure. Eur Heart J. 2023;44(37):3627–39. https://doi.org/10.1093/eurheartj/ehad195.
77. Vaduganathan M, Fonarow GC, Greene SJ, et al. Contemporary treatment patterns and clinical outcomes of comorbid Diabetes Mellitus and HFrEF. JACC: Heart Failure. 2020;8(6):469–80. https://doi.org/10.1016/j.jchf.2019.12.015.
78. Savarese G, Lindenfeld J, Stolfo D, et al. Use of patient-reported outcomes in heart failure: from clinical trials to routine practice. Eur J Heart Fail. 2023;25(2):139–51. https://doi.org/10.1002/ejhf.2778.
79. Rohde MD, French B, Stewart TG, Harrell FE. Bayesian transition models for ordinal longitudinal outcomes. Stat Med. 2024;43(18):3539–61. https://doi.org/10.1002/sim.10133.
80. Pandey A, Kitzman D, Whellan DJ, et al. Frailty among older decompensated heart failure patients. JACC: Heart Fail. 2019;7(12):1079–88. https://doi.org/10.1016/j.jchf.2019.10.003.
81. Pandey A, Kitzman D, Reeves G. Frailty is intertwined with heart failure. JACC: Heart Fail. 2019;7(12):1001–11. https://doi.org/10.1016/j.jchf.2019.10.005.
82. Kokori E, Patel R, Olatunji G, et al. Machine learning in predicting heart failure survival: a review of current models and future prospects. Heart Fail Rev. 2024. https://doi.org/10.1007/s10741-024-10474-y.
83. Kapoor S, Narayanan A. Leakage and the reproducibility crisis in machine-learning-based science. Patterns. 2023;4(9):100804. https://doi.org/10.1016/j.patter.2023.100804.
84. Riley RD, Ensor J, Snell KIE, et al. Calculating the sample size required for developing a clinical prediction model. BMJ. 2020:m441. https://doi.org/10.1136/bmj.m441
85. Snell KIE, Archer L, Ensor J, et al. External validation of clinical prediction models: simulation-based sample size calculations were more reliable than rules-of-thumb. J Clin Epidemiol. 2021;135:79–89. https://doi.org/10.1016/j.jclinepi.2021.02.011.
86. Binuya MAE, Engelhardt EG, Schats W, Schmidt MK, Steyerberg EW. Methodological guidance for the evaluation and updating of clinical prediction models: a systematic review. BMC Med Res Methodol. 2022;22(1):316. https://doi.org/10.1186/s12874-022-01801-8.
87. Riley RD, Pate A, Dhiman P, Archer L, Martin GP, Collins GS. Clinical prediction models and the multiverse of madness. BMC Med. 2023;21(1):502. https://doi.org/10.1186/s12916-023-03212-y.
88. Lala A, Shah P, Khalatbari S, et al. Frailty measures of patient-reported activity and fatigue may predict 1-year outcomes in ambulatory advanced heart failure: a report from the REVIVAL registry. J Card Fail. 2022;28(5):765–74. https://doi.org/10.1016/j.cardfail.2021.10.014.
89. Mirkowski K, Vellone E, Żółkowska B, et al. Frailty and heart failure: clinical insights, patient outcomes and future directions. Card Fail Rev. 2025;11:e05. https://doi.org/10.15420/cfr.2024.34.
90. Rebora P, Scirè CA, Occhino G, et al. Development and validation of an electronic database-based frailty index to predict mortality and hospitalization in a population-based study of adults with SARS-CoV-2. Front Med. 2023;10:1134377. https://doi.org/10.3389/fmed.2023.1134377.
91. Rebora P, Occhino G, Bellelli G, et al. Population-based validation of a frailty index using electronic regional healthcare records for public health use. Sci Rep. 2025;15(1):33298. https://doi.org/10.1038/s41598-025-18611-9.
92. Damen JAAG, Hooft L, Schuit E, et al. Prediction models for cardiovascular disease risk in the general population: systematic review. BMJ. 2016:i2416. https://doi.org/10.1136/bmj.i2416
93. Ahmad FS, Luo Y, Wehbe RM, Thomas JD, Shah SJ. Advances in machine learning approaches to heart failure with preserved ejection fraction. Heart Fail Clin. 2022;18(2):287–300. https://doi.org/10.1016/j.hfc.2021.12.002.
94. Jang SY, Park JJ, Adler E, et al. Mortality prediction in patients with or without heart failure using a machine learning model. JACC: Adv. 2023;2(7):100554. https://doi.org/10.1016/j.jacadv.2023.100554.
95. Tison GH, Sanchez JM, Ballinger B, et al. Passive detection of atrial fibrillation using a commercially available smartwatch. JAMA Cardiol. 2018;3(5):409. https://doi.org/10.1001/jamacardio.2018.0136.
96. Odeh VA, Chen Y, Wang W, Ding X. Recent advances in the wearable devices for monitoring and management of heart failure. Rev Cardiovasc Med. 2024;25(10):386. https://doi.org/10.31083/j.rcm2510386.
97. Spethmann S, Hindricks G, Koehler K, et al. Telemonitoring for chronic heart failure: narrative review of the 20-year journey from concept to standard care in Germany. J Med Internet Res. 2024;26:e63391. https://doi.org/10.2196/63391.
98. Pal M, Parija S, Panda G, Dhama K, Mohapatra RK. Risk prediction of cardiovascular disease using machine learning classifiers. Open Med. 2022;17(1):1100–13. https://doi.org/10.1515/med-2022-0508.
99. Averbuch T, Sullivan K, Sauer A, et al. Applications of artificial intelligence and machine learning in heart failure. Eur Heart J - Digit Health. 2022;3(2):311–22. https://doi.org/10.1093/ehjdh/ztac025.
100. Duraes AR, Gomes-Neto M, Bocchi EA. Editorial: contemporary applications of machine learning and artificial intelligence for the management of heart failure. Front Cardiovasc Med. 2025;12:1611859. https://doi.org/10.3389/fcvm.2025.1611859.
101. Collins GS, Moons KGM. Reporting of artificial intelligence prediction models. Lancet. 2019;393(10181):1577–9. https://doi.org/10.1016/S0140-6736(19)30037-6.
102. Van Der Ploeg T, Austin PC, Steyerberg EW. Modern modelling techniques are data hungry: a simulation study for predicting dichotomous endpoints. BMC Med Res Methodol. 2014;14(1):137. https://doi.org/10.1186/1471-2288-14-137.
103. Tsegaye B, Snell KIE, Archer L, et al. Larger sample sizes are needed when developing a clinical prediction model using machine learning in oncology: methodological systematic review. J Clin Epidemiol. 2025;180:111675. https://doi.org/10.1016/j.jclinepi.2025.111675.

Publisher's Note