

Clinical Prognostic Scoring Systems in Heart Failure with Preserved Ejection Fraction: An Integrative Review of Risk Prediction Models

ANAMARIA DRAGHICI^{1,2,*}, GHEORGHE-ANDREI DAN^{1,3}

¹ “Carol Davila” University of Medicine and Pharmacy, 8 Eroii Sanitari Boulevard, Bucharest, 050474, Romania

² Internal Medicine Department, “Academician Ștefan Milcu” Outpatient Diagnostic and Treatment Medical Center, 26 Grigore Cobalcescu, Bucharest, 010192, Romania

³ Academy of Romanian Scientists, 3 Ilfov Street, Bucharest, 030167, Romania

*Corresponding author

Abstract. Introduction: Risk stratification in heart failure with preserved ejection fraction (HFpEF) remains inconsistent in routine care despite multiple prognostic scores. Key controversies persist. One concerns broad, clinically derived scores versus pathophysiology-grounded tools, including diagnostic frameworks used pragmatically for prognostication. Another concern is the distinction between static baseline risk and dynamic risk states that change across hospitalization and follow-up. **Methods:** We performed a narrative, comparative synthesis of clinician-usable prognostic instruments in HFpEF across conceptual domains and care settings. We extracted study context, endpoints, follow-up, and performance metrics. We emphasized feasibility and clinical interpretability. **Results:** Across studies, discrimination of baseline-only clinical instruments is generally moderate. Performance is often higher for tools that reflect congestion and physiological changes, particularly when reassessed at clinically meaningful time points. Discharge lung ultrasound B-lines, reflecting residual pulmonary congestion, frequently predict early post-discharge events. Immuno-nutritional indices derived from routine laboratory tests provide a complementary prognostic signal, especially in older or recently hospitalized patients. Serial trajectories appear more informative than single measurements. Patient-reported health adds independent prognostic information and supports risk communication. Diagnostic frameworks can stratify risk when constituent data are available, and phenotype distributions are compatible, but transportability and calibration remain recurrent limitations. **Conclusion:** HFpEF prognostication is best supported by a layered, phenotype-aware strategy. This approach integrates global clinical risk, selected biomarkers, discharge congestion assessment, and serial reassessment, rather than relying on a single score.

Keywords: heart failure with preserved ejection fraction, prognosis, prognostic scores, risk stratification, congestion, nutritional status.

What is new? What is important?

- Baseline bedside scores show moderate discrimination in heart failure with preserved ejection fraction (HFpEF) and vary by setting and endpoint.
- Dynamic risk states outperform static risk labels. Discharge lung ultrasound B-lines quantify residual congestion and predict early post-discharge events.
- Low-cost immuno-nutritional indices and patient-reported health add independent prognostic information, especially when tracked serially.
- Diagnostic frameworks can stratify risk when complete components are available, but transportability and calibration remain constraints.

INTRODUCTION

Heart failure with preserved ejection fraction (HFpEF) represents at least half of all heart failure cases and is increasingly regarded as a major clinical and public health concern [1]. The syndrome extends

beyond the preserved left ventricular ejection fraction to encompass interacting myocardial, vascular, and extracardiac pro-inflammatory processes that yield a notably heterogeneous clinical phenotype [2]. Patients frequently present with multiple comorbidities, such as hypertension, diabetes mellitus, obesity, or

chronic kidney disease, that complicate both diagnosis and management [3]. Despite its high prevalence, HFpEF remains an area of significant uncertainty, with high rates of morbidity and mortality that continue to challenge routine care [4].

Accurate prognostic assessment is likely to support a personalized management approach [5]. High-dimensional prognostic models, including machine-learning approaches, can achieve excellent discrimination in controlled datasets, but their bedside use is limited by data requirements [6]. By contrast, clinical risk scores rely on information clinicians already collect, so they are fast, transparent, cost-efficient, and reproducible.

This narrative review was developed to address a persistent gap in HFpEF care. Numerous prognostic scores have been published. Their clinical meaning, however, is often difficult to translate across settings, and the evidence base is heterogeneous. Also, cohorts differ in phenotype mix and disease severity, and endpoints are variably defined.

We therefore focused on instruments that clinicians can realistically compute in routine practice. Scores are treated as complementary perspectives on risk, not as candidates for a single ranking. We also analyze diagnostic frameworks that are commonly used to inform prognosis, even though they were developed for diagnostic evaluation. This pragmatic repurposing appears increasingly supported by observational data [7][8], but its performance remains sensitive to context and component availability. By comparing areas of agreement, contradiction, and uncertainty, we aim to clarify what this literature implies for clinical decisions.

This integrative narrative review aims to summarize clinician-usable prognostic scoring systems evaluated in HFpEF across care settings, highlight head-to-head comparisons where available and report key performance metrics (e.g., C-statistic/AUC and risk gradients), and synthesize practical implications for bedside risk stratification.

METHODOLOGY

This narrative review assesses prognostic risk scores applied to patients with heart failure with preserved ejection fraction (HFpEF). We focused on instruments intended for bedside use rather than complex prognostic models or machine-learning approaches.

SEARCH STRATEGY AND STUDY IDENTIFICATION

A focused search was performed in PubMed for human studies in adults (≥ 18 years), restricted to English-language publications from the last 10 years. We used a score-oriented strategy combining HFpEF terminology with both generic score terms and commonly cited score names. We excluded studies that were non-HFpEF-specific, not bedside-applicable, or based on biomarker-only or high-dimensional prediction models.

DATA EXTRACTION AND OUTCOMES OF INTEREST

From each included study, we extracted the study setting and population, the HFpEF definition used, the inputs and scoring method for the score, the endpoints assessed, the follow-up duration when available, and the reported measures of prognostic performance. Endpoints of interest included all-cause death, cardiovascular death, heart-failure hospitalization, and prespecified composite cardiovascular outcomes.

COMPARATIVE FRAMEWORK AND STATISTICAL METRICS

Because the included studies differed in cohort characteristics, endpoint definitions, and follow-up, we did not pool estimates. Instead, we used a structured narrative synthesis to compare instruments across settings and endpoints. Discrimination was captured using AUC or concordance statistics (C-statistic/c-index). Strength of association was summarized using HRs or ORs, retaining the original scaling. Calibration was recorded when reported.

Findings were interpreted by score domain, study design (derivation vs validation), cohort setting, endpoint/time horizon, and feasibility.

REPORTING QUALITY GUIDANCE AND USE OF GENERATIVE AI TOOLS

To support methodological transparency and narrative quality, reporting was guided by the Scale for the Assessment of Narrative Review Articles (SANRA).

ChatGPT (OpenAI) was used as a language-support tool for English translation, grammar correction, and refinement of academic phrasing. Study selection, data extraction, and interpretation were conducted by the authors, who assume full responsibility for the accuracy and scientific integrity of the final manuscript.

OVERVIEW OF INCLUDED SCORES AND COMPONENT DOMAINS

Before detailing individual instruments, Table 1 maps the clinician-usable instruments included in this review. It summarizes each tool's conceptual target and core components, alongside the clinical context in which it has been evaluated in HFpEF, such as acute admission triage, early post-discharge vulnerability, longitudinal ambulatory follow-up, or phenotype-specific pathways. The table also flags an important distinction for interpretation: several widely used scores were derived in mixed heart-failure cohorts and only later tested in HFpEF subsets, whereas other tools were developed or validated more directly in HFpEF-oriented populations.

This difference has implications for external validity when results are applied across settings.

Because HFpEF case definitions varied across included cohorts (e.g., LVEF thresholds, use of natriuretic peptides, and diagnostic frameworks), these definitions are summarized in Supplementary Table 1 to aid interpretation of cross-study comparisons.

Supplementary Table 2 and 3 summarize outcome-oriented performance (mortality and hospitalization, respectively), whereas Table 2 synthesizes comparative applicability across clinical settings and phenotypes, highlighting when a tool is most informative and which inputs constrain bedside feasibility.

Table 1

Main characteristics of prognostic scores

Score (acronym)	Expanded name	Core components (concise)	Conceptual domain	Typical setting / use
ARIC AD-HFpEF [11]	ARIC Acute Decompensated HFpEF score	Age; systolic BP; BUN; sodium; hypoxia; heart rate; natriuretic peptides; anemia; underweight	Acute clinical burden / triage	Acute HFpEF admission; 28-day/1-year risk estimate
C2HEST [20]	Coronary artery disease; COPD; Hypertension; Elderly; Systolic HF; Thyroid disease	CAD; COPD; hypertension; age ≥ 75 ; prior systolic HF; thyroid disease	Comorbidity / systemic burden	Ambulatory HFpEF (TOPCAT) for background risk enrichment
CONUT [21][23][26][27][28]	Controlling Nutritional Status	Albumin; total cholesterol; lymphocyte count	Nutrition–inflammation	Older / hospitalized or post-discharge HFpEF
GNRI [22][24][27][28]	Geriatric Nutritional Risk Index	Albumin; weight-to-ideal-weight term	Nutrition–inflammation / frailty	Older/frail; hospitalized or ambulatory HFpEF
GWTHG-HF [10]	Get With The Guidelines–Heart Failure risk score	Age; SBP; BUN; sodium; heart rate; COPD; race	Global clinical risk (mixed-EF derivation)	In-hospital acute HF; post-discharge risk stratification
H2FPEF [13][14][15][16][17]	Heavy; Hypertensive; Atrial fibrillation; Pulmonary hypertension; Elder; Filling pressure	BMI ≥ 30 ; ≥ 2 antihypertensives; AF; PASP/PH; age > 60 ; E/e'	Diagnostic framework repurposed for prognosis	Ambulatory/inpatient HFpEF; useful when stress testing unavailable
HFA-PEFF [14][15][17]	Heart Failure Association Pre-test assessment, Echocardiography & NP, Functional testing, Final aetiology	Echo + NP domains; functional testing (step-3) when available	Diagnostic framework repurposed for prognosis	Suspected/confirmed HFpEF; prognostic value highest with complete work-up
KCCQ [33]	Kansas City Cardiomyopathy Questionnaire (incl. KCCQ-12)	Patient-reported symptoms, function, quality of life	Patient-reported health status	Ambulatory/chronic HF (incl. HFpEF subsets) for outcome prediction and communication

Table 1 (continued)

LUS B-lines [19]	Lung ultrasound B-line count	B-line burden at discharge (residual pulmonary congestion)	Congestion physiology	Discharge risk stratification after acute HF (including HFpEF)
MAGGIC [9][17]	Meta-Analysis Global Group in Chronic HF risk score	Demographics; clinical status; comorbidities; therapies (standard MAGGIC variables)	Global clinical risk (mixed-EF derivation)	Chronic HF; validated in HFpEF subsets; longitudinal risk
MEDIA [18]	MEDIA echocardiographic score	PASP >40; IVC collapsibility <50%; average E/e' >9; lateral s' <7	Echocardiography- only hemodynamic burden	Acute HFpEF and stable outpatient HFpEF
mGPS [29]	Modified Glasgow Prognostic Score	C-reactive protein; albumin	Inflammation– nutrition	Ambulatory HFpEF; 12-month outcomes
NRS-2002 [30]	Nutritional Risk Screening 2002	BMI; weight loss; intake reduction; disease severity	Bedside nutrition screening	Acute HFpEF admissions; in-hospital risk (sex-specific effects)
SHFM [9]	Seattle Heart Failure Model	Clinical variables; labs; therapies; device therapy (model inputs)	Global clinical risk (mixed-EF derivation)	Chronic HF; comparator in HFpEF validation work
TRI [12]	TIMI Risk Index	Heart rate $\times$ (age/10) ² $\div$ systolic BP	Ultra-parsimonious acute triage	Acute HFpEF admission; in-hospital mortality triage
HALO [31]	HFpEF survivAL hOspitalization (HALO) score	Clinical severity; echocardiographic burden; natriuretic peptide load; prior HF hospitalization count	Multimodal HFpEF event risk (survival + recurrent admissions)	Recently hospitalized HFpEF; post-discharge risk stratification for survival and future admission burden
WATCH-DM [32]	WATCH-DM risk score	Age; BMI; BP; fasting glucose; creatinine; HDL-C; QRS; prior MI/CABG	Phenotype-specific (diabetes)	T2DM + HFpEF at discharge; ~1-year mortality

Abbreviations: AF – atrial fibrillation; BMI – body mass index; BNP – B-type natriuretic peptide; BUN – blood urea nitrogen; CONUT – Controlling Nutritional Status; CRP – C-reactive protein; GNRI – Geriatric Nutritional Risk Index; HALO – HFpEF survivAL hOspitalization; HFpEF – heart failure with preserved ejection fraction; IVC – inferior vena cava; KCCQ – Kansas City Cardiomyopathy Questionnaire; LUS – lung ultrasound; LVEF – left ventricular ejection fraction; mGPS – modified Glasgow Prognostic Score; MLHFQ – Minnesota Living with Heart Failure Questionnaire; NT-proBNP – N-terminal pro-B-type natriuretic peptide; PASP – pulmonary artery systolic pressure; PNI – Prognostic Nutritional Index; TRI – TIMI Risk Index.

GLOBAL CLINICAL RISK SCORES USABLE AT THE BEDSIDE

MAGGIC (Meta-Analysis Global Group in Chronic Heart Failure) and GWTG-HF (Get With The Guidelines–Heart Failure) remain common entry points because they can be calculated quickly, using variables that are usually available early in care. Both tools were derived in broader, mixed-ejection fraction heart-failure populations. Most HFpEF evidence therefore reflects subgroup validation, not HFpEF-only derivation. This may partly explain why discrimination in HFpEF tends to land in the moderate range, with calibration and

risk separation varying across settings and endpoints (see Supplementary Table 2 and Supplementary Table 3) [9][10]. MAGGIC appears to retain prognostic value in HFpEF, particularly once natriuretic peptide information is incorporated, although the incremental benefit may depend on cohort severity mix and endpoint definition (all-cause vs HF-specific hospitalization; first vs recurrent events) [9]. Where head-to-head data are available, MAGGIC performs similarly to other established global models in HFpEF. In the Rich *et al.* [9] validation cohort, discrimination for all-cause death was MAGGIC C-statistic 0.74 (95% CI 0.68–0.80) versus SHFM 0.72 (95% CI

0.67–0.78), with BNP alone performing comparably (0.76; 95% CI 0.70–0.81).

GWTG-HF, developed for in-hospital risk stratification in acute heart failure, appears to carry forward clinically meaningful stratification after discharge in HFpEF, but reported performance varies with acuity at index admission, comorbidity burden, and local discharge practices [10].

For acute triage, the ARIC AD-HFpEF (Atherosclerosis Risk in Communities Acute Decompensated HFpEF) score is conceptually distinct because it was designed around HFpEF admissions and short-to-intermediate follow-up. It formalizes what many clinicians already use implicitly at presentation, such as blood pressure, renal indices, oxygenation, natriuretic peptides, and markers of frailty [11]. The TIMI Risk Index (TRI) pushes simplicity even further. It uses only heart rate, age, and systolic blood pressure and still appears to identify a higher in-hospital mortality risk in acute HFpEF, making it attractive when time and data are limited [12]. The trade-off is predictability. TRI is best suited for front-end triage, not longitudinal risk (Supplementary Table 2) [12].

DIAGNOSTIC SCORES REPURPOSED FOR PROGNOSIS

Diagnostic frameworks in HFpEF are increasingly used to estimate prognosis, although they were designed for case identification rather than outcome prediction. H2FPEF and HFA-PEFF are primarily diagnostic decision-support tools. Therefore, their prognostic use in established HFpEF should be considered off-label and interpreted with caution. This repurposing is still understandable. These scores capture pathophysiology linked to HFpEF events and can be calculated from routine clinical data collected during evaluation. Across ambulatory and inpatient cohorts, H2FPEF (Heavy, Hypertensive, Atrial Fibrillation, Pulmonary Hypertension, Elder, Filling pressure) generally stratifies risk of death and heart-failure hospitalization, and higher scores align with worse physiology, including exercise limitation and adverse hemodynamics [13][14]. However, cohorts differ in atrial fibrillation prevalence, obesity distribution, and diagnostic work-up. As a result, score distributions and observed risk gradients may shift even when the same cut-points are applied.

Head-to-head comparisons often favor H2FPEF over the HFA-PEFF algorithm when step-3 (stress) testing is unavailable, likely because H2FPEF

can be computed from routine clinical data and resting echocardiography alone [14][15]. HFA-PEFF comprises Heart Failure Association-led pre-test assessment, echocardiography and natriuretic peptides, functional testing, and final etiologic classification. As summarized in Table 2, the apparent prognostic separation of these diagnostic frameworks is strongly conditioned by feasibility. Still, HFA-PEFF has prognostic value when natriuretic peptide data and stress-echocardiography components are obtained. Practical barriers and variable discrimination across settings may limit its use as a stand-alone tool [15]. In a hospitalized HFpEF cohort, the H2FPEF score correlated inversely with peak left-atrial strain and identified a higher risk of death or HF hospitalization, suggesting that H2FPEF continues to track atrial remodeling and carries prognostic information when stress testing is not feasible [16]. Przewłocka-Kośmala *et al.* [17] studied patients with exertional dyspnea and suspected HFpEF. In that head-to-head comparison, discrimination was similar for MAGGIC, H2FPEF, and step-2 HFA-PEFF (Harrell's $C = 0.637, 0.644, \text{ and } 0.638$, respectively), whereas incorporating step-3 exercise data improved discrimination (Harrell's $C = 0.715$) [17]. This finding is setting-dependent (suspected HFpEF physiology cohort) and constrained by eligibility, because only patients in sinus rhythm were included (see Supplementary Table 2 and Supplementary Table 3).

IMAGING AND FUNCTIONAL CONGESTION METRICS

Echo-anchored tools are often appealing in HFpEF because they target filling pressures, pulmonary hypertension, and systemic congestion, which are processes that are tightly linked to decompensation and subsequent hospital admission. The MEDIA score, based on pulmonary artery systolic pressure, inferior vena cava collapsibility, E/e' , and lateral s' , offers a compact way to translate a multiparametric echocardiogram into a structured risk estimate [18]. Its reported performance differs by setting. In acute HFpEF, higher scores are more strongly associated with short-term mortality. In stable outpatients, the signal shifts toward future heart-failure hospitalization (Supplementary Table 2 and Supplementary Table 3) [18]. This pattern suggests setting-dependent prognostic meaning.

Because HFpEF events are closely linked to congestion, markers of residual hemodynamic burden at discharge carry prognostic weight. Lung

ultrasound (LUS) B-lines quantify pulmonary congestion. In acute HFpEF, a high B-line burden at discharge predicts 6-month rehospitalization or death and, in several analyses, outperforms admission measures [19]. These results come primarily from early post-discharge windows and discharge-based protocols, so they should not be compared directly with admission-based scores or longer follow-up cohorts. Combining discharge LUS with BNP (B-type natriuretic peptide) and diastolic parameters

further improves classification [19]. A simple B-line count (using a Youden cut-point of 22 B-lines) can function as an imaging-based risk metric that complements global scores (e.g., GWTG-HF, MAGGIC) and HFpEF-specific instruments (MEDIA) with utility in the vulnerable post-discharge window. LUS behaves more like a modifiable risk state than a static label, which may explain its incremental value when measured at discharge rather than at presentation (Table 2).

Table 2

Practical applicability of clinician-usable HFpEF prognostic tools across settings and phenotypes

Clinical context	Most feasible instruments	What they primarily capture	When they tend to add value	Key limitations / cautions
Acute admission triage (HFpEF)	ARIC AD-HFpEF; GWTG-HF; TRI [10][11][12]	Short-term mortality risk using readily available vitals and labs	Early triage and discharge planning; prioritizing follow-up intensity	GWTG-HF derived in mixed-EF populations; TRI trades physiologic specificity for speed
Discharge vulnerability after acute HF	LUS B-lines ($\pm$ BNP/echo) [19]	Residual pulmonary congestion	Near-term readmission/death risk; decongestion targets before discharge	Requires operator familiarity; cut-points may vary across protocols
Ambulatory / longitudinal risk	MAGGIC ($\pm$ BNP); KCCQ [9][33]	Global clinical burden and patient-reported status	Longitudinal counselling and shared decisions; complements event-focused tools	MAGGIC derived in mixed-EF cohorts; KCCQ evidence often from mixed HF samples
Echocardiography-only hemodynamic burden	MEDIA; H2FPEF (resting echo) [16][18]	Filling pressure, pulmonary pressure, venous congestion, longitudinal function	When natriuretic peptides are equivocal or stress testing unavailable	Generalizability depends on echo quality and cohort characteristics (e.g., AF exclusion in MEDIA ambulatory validation)
Older / frail or recently hospitalized HFpEF	GNRI; CONUT; mGPS; NRS-2002 [21][22][23][24][29][30]	Frailty-like vulnerability, malnutrition, inflammation	Risk refinement beyond BMI; identifies patients who may benefit from nutritional evaluation and closer follow-up	Definitions and thresholds differ; readmission associations may be less consistent than mortality
Phenotype-specific pathways	WATCH-DM (T2DM); C2HEST (comorbidity burden); nutrition indices for rhythm pathway [20][25][32]	Diabetes burden, systemic comorbidity, vulnerability affecting rhythm and recovery	Discharge decisions and surveillance tailored to phenotype	Derived/validated in specific cohorts; may require local adaptation to preserve performance
Multimodal HFpEF event risk	HALO; prior HF hospitalization count [31]	Structural/hemodynamic load plus prior events	Higher-risk recurrent admission phenotype	Data requirements (echo + biomarkers + history); external validation still limited

Abbreviations: AF – atrial fibrillation; HFH – heart failure hospitalization; LUS – lung ultrasound; NP – natriuretic peptides.

PRAGMATIC RISK TOOLS OUTSIDE HFPEF-SPECIFIC SCORES

Some instruments enter HFpEF practice from adjacent clinical problems. The C2HEST score, initially developed to estimate incident atrial fibrillation, illustrates how a simple comorbidity-weighted framework may still enrich baseline risk in HFpEF, with associations extending beyond rhythm outcomes to death and hospitalization [20]. This should not be overinterpreted as a superior HFpEF prognostic model. Instead, it may be more appropriately interpreted as a low-friction baseline risk layer that reflects age and systemic comorbidity burden, factors that remain relevant across HFpEF phenotypes regardless of the dominant HFpEF mechanism (Supplementary Table 3) [20].

IMMUNO-NUTRITIONAL SCORES FOR PROGNOSIS

A frequent topic across contemporary cohorts is that malnutrition and systemic inflammation carry independent prognostic weight in HFpEF and can be captured with low-cost indices derived from routine laboratory testing. GNRI (Geriatric Nutritional Risk Index), CONUT (Controlling Nutritional Status), and PNI (Prognostic Nutritional Index) each associate with mortality and with heart-failure readmission across several datasets (Supplementary Table 2 and 3) [21][22][23]. Among these measures, CONUT and GNRI tend to be most informative in older, hospitalized, or recently discharged cohorts [22][23]. Outpatient cohorts often show weaker absolute gradients, and follow-up horizons vary, limiting direct comparisons (Table 2).

In TOPCAT-Americas (Treatment of Preserved Cardiac Function Heart Failure with an Aldosterone Antagonist-American cohort), roughly one-third of participants met GNRI criteria for nutritional risk, and lower GNRI values independently tracked higher cardiovascular events and mortality [24]. This is trial-derived evidence in a selected population, not a universal HFpEF registry estimate. Nutritional risk quantified by CONUT, PNI, and NRI identified patients at higher risk of adverse rhythm-related outcomes in an HFpEF cohort undergoing rhythm control, reinforcing the broader prognostic role of malnutrition beyond mortality and HF readmission [25]. That cohort differs by pathway (rhythm management), endpoint mix, and comorbidity profile, so results should be interpreted within that context.

Equally important seems to be the temporal element. Worsening nutritional trajectories after discharge or during the index hospitalization (e.g., rising CONUT or falling GNRI) are associated with a higher subsequent risk, whereas improvement in these indices is associated with fewer later events (Supplementary Table 3) [26][27][28].

The modified Glasgow Prognostic Score (mGPS), which integrates CRP (C-reactive protein) and albumin, appears to add value over NT-proBNP for predicting 12-month death or HF hospitalization in ambulatory patients with HFpEF (Supplementary Table 2 and 3), thereby linking inflammation to clinically relevant outcomes [29]. Bedside instruments such as the nutritional index NRS-2002 (Nutritional Risk Screening 2002) can further identify high-risk inpatients, particularly men. This finding implies a sex-specific vulnerability that can be easily integrated into routine practice during hospitalization without additional cost [30]. From an implementation perspective, these indices are inexpensive and reproducible.

BIOMARKER-ENRICHED AND MULTIMODAL HFPEF-SPECIFIC TOOLS

Scores that combine biomarkers, cardiac structure, and prior clinical events reflect the idea that HFpEF risk has multiple drivers, so focusing on a single type of information is unlikely to capture the full risk profile. HALO is a clear example. By integrating clinical severity, echocardiographic burden, natriuretic peptide load, and prior hospitalizations, it predicts survival and future admission burden, treating recurrent decompensation as part of the disease trajectory rather than a downstream complication [31]. WATCH-DM, oriented to the diabetic phenotype, uses routinely available discharge variables and appears to stratify 1-year mortality after HFpEF hospitalization, with performance comparable to MAGGIC, while being more explicitly tailored to a cardiometabolic risk profile [32]. Both tools are most interpretable when applied to the populations they were tested in (post-hospitalization cohorts; diabetes enrichment for WATCH-DM). They are less suited to uniform application across all HFpEF presentations (see comparative notes in the tables) [31][32].

Patient-reported health adds a complementary dimension. Kansas City Cardiomyopathy Questionnaire (KCCQ or KCCQ-12) predicts death and hospitalization (Supplementary Table 2 and 3). In a direct comparison, KCCQ showed higher

discrimination than the Minnesota Living With Heart Failure Questionnaire (MLHFQ) (C-index 0.702 [95% CI 0.666–0.738] vs 0.658 [95% CI 0.621–0.695]) and a stronger risk gradient per 5-point change in the expected direction (adjusted HR 0.894 per 5-point KCCQ increase vs 1.077 per 5-point MLHFQ increase) [33]. Higher KCCQ reflects better health status, whereas higher MLHFQ reflects worse status. However, studies vary by instrument version, follow-up, and endpoint definition, which should be considered when comparing effect sizes across cohorts.

TEMPORAL TRAJECTORIES AND PHENOTYPE-AWARE APPLICATION

What happens over time tends to matter more than a single snapshot. Worsening CONUT or declining GNRI over months or even within a single hospitalization independently signals a higher subsequent risk, whereas stabilization or improvement appears to attenuate it [26][27][28]. Natriuretic peptide trajectories show similar dynamics and are partly captured by HALO via prior admission count [31]. The practical implication is that risk assessment should move beyond one-time scoring. Reevaluate with low-cost indices (CONUT, GNRI), global tools (MAGGIC, GWTG-HF), and congestion metrics (LUS B-lines) at clinically meaningful checkpoints, such as admission, discharge, 30–90 days, and after any re-hospitalization [28] [19]. Such serial assessments are straightforward and likely to catch turning points that a baseline score misses.

Building on this emphasis on reassessment, a layered workflow can align tools with the decision point. At admission or discharge, a global score can provide baseline context for triage and early planning (e.g., MAGGIC or GWTG-HF, and ARIC/TRI in acute presentations) [9][10][11][12]. At discharge, adding a congestion-focused measure such as LUS B-lines helps quantify residual risk that baseline clinical variables may not capture [19]. During early follow-up (around 30–90 days) and after any rehospitalization, repeating low-cost nutrition–inflammation indices (CONUT, GNRI) and updating biomarker burden can track whether risk is stable or undergoing meaningful change (see Figure 1) [26][27][28]. Across cohorts, congestion and trajectory sensitive measures tend to outperform baseline-only tools for near-term events.

The choice and sequencing of layers then become phenotype dependent. In older or frail

patients, nutrition-based scores (GNRI, CONUT, mGPS) often outperform anthropometrics such as BMI, which is fluid-sensitive and may miss sarcopenia. These findings suggest that nutritional frailty is likely a key driver of events in this population [21][22] [23] [29]. In phenotype-specific settings, nutritional indices also inform prognosis: higher CONUT and lower PNI/NRI were associated with a greater risk of adverse rhythm outcomes in HFpEF patients undergoing rhythm control, supporting the use of nutrition screening to guide follow-up frequency and intervention thresholds [25]. In the diabetic phenotype, WATCH-DM provides a straightforward, diabetes-tailored stratification tool at discharge. Used alongside NT-proBNP (N-terminal pro-B-type natriuretic peptide) and, where feasible, KCCQ, it synthesizes biological signals with patient-reported outcomes [32][33]. Sex differences warrant attention as well. Bedside screening with NRS-2002 uncovers a pronounced mortality indicator in men during acute admissions, arguing for sex-aware escalation of nutritional support [30]. Finally, generalizability is not guaranteed. When phenotype distributions differ from derivation cohorts, local recalibration is preferable [13].

LIMITATIONS

This review is narrative and integrative rather than a preregistered systematic review, and we did not pool estimates. The main limitation is heterogeneity across the underlying studies. Cohorts differed by setting (hospitalized, post-discharge, outpatient, trial), HFpEF definitions and diagnostic work-up, phenotype distributions, and endpoints. As a result, numerical performance metrics are not fully comparable, and transportability across contexts is constrained.

Reporting was inconsistent. Discrimination was frequently provided, whereas calibration was variably assessed and rarely standardized. Effect measures were reported on different scales, and robust external validation and within-cohort head-to-head comparisons were limited.

In addition, several tools discussed as prognostic tools (e.g., H2FPEF and HFA-PEFF) were originally designed for diagnostic evaluation; therefore, prognostic use represents off-label repurposing and may be sensitive to missing components and local phenotype mix.

Finally, the focus on clinician-usable scores may underrepresent higher-dimensional models that

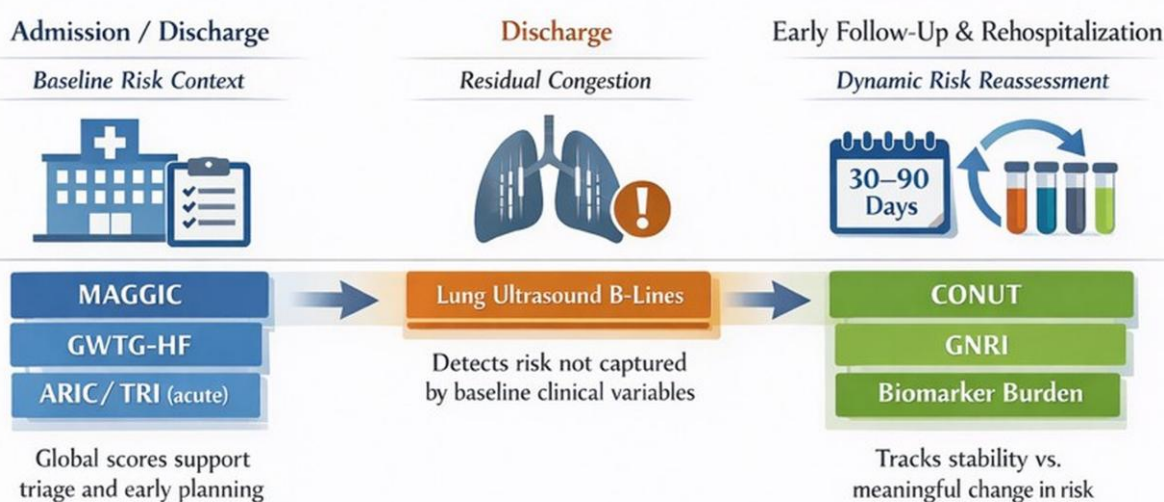
could improve prediction but require infrastructure and recalibration before routine implementation.

CONCLUSIONS

Baseline clinical scores in HFpEF offer only moderate discrimination. Performance appears to improve when assessment reflects what drives events, such as residual congestion, nutritional or inflammatory state, and how these change over time. A layered workflow seems most practical: start with an implementable clinical or diagnostic

score (MAGGIC, GWTG-HF, H2FPEF/HFA-PEFF), add natriuretic peptides, check discharge lung ultrasound for B-lines, and follow simple nutrition indices plus KCCQ longitudinally. This combination is feasible, low-cost, and suitable for routine clinical use. It may support admission triage, early post-discharge planning, and follow-up. Utility is likely phenotype-dependent: nutrition-centric indices are particularly informative in older or frail patients, while WATCH-DM provides diabetes-specific stratification.

Layered Risk Reassessment Aligned to Clinical Decision Points



Risk in heart failure is dynamic—there is a need to move beyond a one-time score!

Figure 1. Layered risk reassessment aligned to clinical decision points.

Introducere: Stratificarea riscului în insuficiența cardiacă cu fracție de ejeție păstrată (HFpEF) rămâne neuniformă în practica curentă, în ciuda existenței multiplelor scoruri prognostice. Persistă controverse importante. Una dintre acestea privește utilizarea scorurilor largi, derivate clinic, comparativ cu instrumentele fundamentate pe mecanisme fiziopatologice, inclusiv cadrele diagnostice utilizate pragmatic în scop prognostic. O altă controversă vizează distincția dintre riscul static, evaluat la momentul inițial, și stările de risc dinamice, care se modifică pe parcursul spitalizării și în perioada de urmărire.

Metode: Am realizat o sinteză narativă comparativă a instrumentelor prognostice utilizabile în practica clinică la pacienții cu HFpEF, analizate în funcție de domeniul conceptual și contextul de îngrijire. Au fost extrase date privind contextul studiilor, criteriile de evaluare (endpoints), durata urmăririi și indicatorii de performanță. Am pus accent pe fezabilitatea utilizării și interpretabilitatea clinică.

Rezultate: În majoritatea studiilor, capacitatea de discriminare a instrumentelor clinice bazate exclusiv pe evaluarea inițială este, în general, moderată. Performanța este adesea superioară în cazul instrumentelor care reflectă congestia și modificările fiziologice, în special atunci când sunt reevaluate la momente clinice relevante. Liniile B evidențiate prin

ecografie pulmonară la externare, reflectând congestia pulmonară reziduală, prezic frecvent evenimentele precoce post-externare. Indicii imuno-nutriționali derivați din analize de laborator uzuale oferă un semnal prognostic complementar, în special la pacienții vârstnici sau recent spitalizați. Traectoriile seriate par a fi mai informative decât determinările unice. Evaluarea stării de sănătate raportată de pacient aduce informații prognostice independente și sprijină comunicarea riscului. Cadrele diagnostice pot stratifica riscul atunci când datele componente sunt disponibile și distribuțiile fenotipice sunt compatibile, însă transportabilitatea și calibrarea rămân limitări recurente.

Concluzii: Evaluarea prognosticului în HFpEF este optim susținută de o strategie stratificată, adaptată fenotipului. Această abordare integrează riscul clinic global, biomarkeri selectați, evaluarea congestiei la externare și reevaluarea serială, în locul utilizării unui singur scor prognostic.

Correspondence to: Anamaria Draghici, M.D.

Internal Medicine Department, “Academician Ștefan Milcu” Outpatient Diagnostic and Treatment Medical Center, 26 Grigore Cobalcescu, 010192, Sector 1, Bucharest, Romania
Tel: +40720804638
Email: anamaria_tocitu@yahoo.com

Grant support/Funding: None.

Declaration of interest: The authors declare that there are no conflicts of interest.

REFERENCES

1. KITTLESON M., PANJRATH G., AMANCHERLA K., DAVIS L., DESWAL A., DIXON D., *et al.* 2023 ACC Expert Consensus Decision Pathway on Management of Heart Failure With Preserved Ejection Fraction: A Report of the American College of Cardiology Solution Set Oversight Committee. *Am Coll Cardiol.* 2023;**81**(18):1835–1878.
2. MA C., LUO H., FAN L., LIU X., GAO C. *Heart failure with preserved ejection fraction: an update on pathophysiology, diagnosis, treatment, and prognosis.* *Braz J Med Biol Res.* 2020;**53**(7):e9646.
3. DEICHL A., WACHTER R., EDELMANN F. *Comorbidities in heart failure with preserved ejection fraction.* *Herz.* 2022;**47**(4):301–307.
4. SHAHIM A., HOURQUEIG M., LUND L.H., SAVARESE G., OGER E., VENKATESHVARAN A., *et al.* Long-term outcomes in heart failure with preserved ejection fraction: Predictors of cardiac and non-cardiac mortality. *ESC Hear. Fail.* 2023;**10**(3):1835–1846.
5. SKOULARIGKIS S., KOUREK C., XANTHOPOULOS A., BRIASOULIS A., ANDROUTSOPOULOU V., MAGOULIOTIS D., *et al.* Prognostic Models in Heart Failure: Hope or Hype? *J Pers Med.* 2025;**15**(8):345.
6. JIA Y., CUI N., JIA T., SONG J. *Prognostic models for patients suffering a heart failure with a preserved ejection fraction : a systematic review.* *ESC Hear. Fail.* 2024;**11**(3):1341–1351.
7. JORGE A.J.L., ROSA M.L.G., MARTINS W.D.A., LEITE A., CORREIA D.M. da S., SAAD M.A.N., *et al.* Prognosis of Heart Failure with Preserved Ejection Fraction in Primary Care by the H2FPEF Score. *Int. J. Cardiovasc. Sci.* 2020;**33**(6):666–672.
8. EGASHIRA K., SUETA D., KOMORITA T., YAMAMOTO E., USUKU H., TOKITSU T., *et al.* HFA-PEFF scores: prognostic value in heart failure with preserved left ventricular ejection fraction. 2022; *Korean J Intern Med.* 2022;**37**:96–108.
9. RICH J.D., BURNS J., FREED B.H., MAURER M.S., BURKHOFF D., SHAH S.J. *Meta-Analysis Global Group in Chronic (MAGGIC) Heart Failure Risk Score: Validation of a Simple Tool for the Prediction of Morbidity and Mortality in Heart Failure With Preserved Ejection Fraction.* *J Am Hear. Assoc.* 2018;**7**(20):e009594.
10. SUZUKI S., YOSHIHISA A., SATO Y., KANNO Y., WATANABE S., ABE S., *et al.* Clinical Significance of Get With the Guidelines-Heart Failure Risk Score in Patients With Chronic Heart Failure After Hospitalization. *J Am Hear. Assoc.* 2018;**7**(17):e008316.
11. THORVALDSEN T., CLAGGETT B.L., SHAH A., CHENG S., AGARWAL S.K., WRUCK L.M., *et al.* Predicting risk in patients hospitalized for acute decompensated heart failure and preserved ejection fraction: The Atherosclerosis Risk in Communities Study Heart Failure Community Surveillance. *Circ Hear. Fail.* 2017;**10**(12):e003992.
12. MATSUSHITA K., HARADA K., KOHNO T., NAKANO H., KITANO D., TAKEI M., *et al.* A Simple Prognostic Index Based on Heart Rate, Age, and Systolic Blood Pressure for the Risk Stratification of Heart Failure With Preserved Ejection Fraction. *Hear. Lung Circ.* 2025;**34**(9):883–890.
13. TAKAHARI K., HIDAKA T., UEDA Y., IZUMI K., HARADA Y., SUSAWA H., *et al.* H2FPEF Score for the Prediction of Exercise Intolerance and Abnormal Hemodynamics in Japanese – Evaluation by Exercise Stress Echocardiography Combined With Cardiopulmonary Exercise Testing. *Circ J.* 2019;**83**(12):2487–2493.

14. AMANAI S., HARADA T., KAGAMI K., YOSHIDA K., KATO T. *The – H₂ FPEF and HFA – PEFF algorithms for predicting exercise intolerance and abnormal hemodynamics in heart failure with preserved ejection fraction*. Sci. Rep. 2022;**12**(1):13.
15. BARROS F.C., CEZARO J.C. De, COSTA P.D., COSTA G.D., BARRETO A., SANTOS S., et al. *Prognostic Applications of Current Clinical Scores in Heart Failure with Preserved Ejection Fraction: A Prospective Cohort Study*. Arq Bras Cardiol. 2025;**122**(6):e20240852.
16. HWANG I., CHO G., CHOI H., YOON Y.E., PARK J.J., PARK J., et al. *H₂FPEF Score Reflects the Left Atrial Strain and Predicts Prognosis in Patients With Heart Failure With Preserved Ejection Fraction*. J. Card. Fail. 2020;**27**(2):198–207.
17. PRZEWOŁOCKA-KOSMALA M., BUTLER J., DONAL E., PONIKOWSKI P., KOSMALA W. *Prognostic Value of the MAGGIC Score, H₂FPEF Score, and HFA-PEFF Algorithm in Patients with Exertional Dyspnea and the Incremental Value of Exercise Echocardiography*. J Am Soc Echocardiogr. 2022;**35**(9):966–975.
18. COIRO S., HUTTIN O., KOBAYASHI M., LAMIRAL Z., ZANNAD F., ROSSIGNOL P., et al. *Validation of the MEDIA echo score for the prognosis of heart failure with preserved ejection fraction*. Hear. Fail Rev. 2023;**28**(2):453–464.
19. PALAZZUOLI A., RUOCCO G., BELTRAMI M., NUTI R., CLELAND J.G. *Combined use of lung ultrasound, B-type natriuretic peptide, and echocardiography for outcome prediction in patients with acute HFrEF and HFpEF*. Clin. Res. Cardiol. 2018;**107**(7):586–596.
20. LIANG W., WU Y., XUE R., WU Z., WU D., HE J., et al. *C₂ HEST score predicts clinical outcomes in heart failure with preserved ejection fraction: a secondary analysis of the TOPCAT trial*. BMC Med. 2021;**19**(1):44.
21. CHIEN S., LO C., LIN C., SUNG K., TSAI J., HUANG W., et al. *Malnutrition in acute heart failure with preserved ejection fraction: clinical correlates and prognostic implications*. ESC Hear. Fail. 2019;**6**(5):953–964.
22. NISHI I., SEO Y., HAMADA-HARIMURA Y., YAMAMOTO M., ISHIZU T., SUGANO A., et al. *Geriatric nutritional risk index predicts all-cause deaths in heart failure with preserved ejection fraction*. ESC Hear. Fail. 2019;**6**(2):396–405.
23. CHEN Y., ZHENG H., HE Y. *Prognostic significance of controlling nutritional status in older adults with heart failure with preserved ejection fraction: a prospective comparative study with other objective nutritional indices*. Aging Clin Exp Res. 2023;**35**(6):1305–1315.
24. MINAMISAWA M., SEIDELMANN S.B., CLAGGETT B., HEGDE S.M., SHAH A.M., DESAI A.S., et al. *Impact of Malnutrition Using Geriatric Nutritional Risk Index in Heart Failure With Preserved Ejection Fraction*. JACC Hear. Fail. 2019;**7**(8):664–675.
25. ZHANG Z., WANG C., LIU Q., XIAO Y., ZHOU J., WU K., et al. *The risk of malnutrition as a predictor of arrhythmia recurrence after catheter ablation in patients with paroxysmal non-valvular atrial Fibrillation and heart failure with preserved ejection fraction*. PLoS One. 2025;**20**(1):e0317721.
26. KITAO T., HIKOSO S., TAMAKI S., SEO M., YANO M., HAYASHI T., et al. *Post-discharge changes in nutritional status predict prognosis in patients with acute decompensated HFpEF from the PURSUIT – HFpEF Registry*. Hear. Vessel. 2025;**40**(7):577–591.
27. SUNAGA A., HIKOSO S., YAMADA T., YASUMURA Y., TAMAKI S., KURAKAMI H., et al. *Change in Nutritional Status during Hospitalization and Prognosis in Patients with Heart Failure with Preserved Ejection Fraction*. Nutrients. 2022;**14**(20):4345.
28. LU H., CLAGGETT B., MINAMISAWA M., OSTROMINSKI J., FOA A., PABÓN M., et al. *Prognostic Significance of Nutritional Scores in Patients With Heart Failure: Insights From the PARAGON- HF Trial*. J Am Hear. Assoc. 2025;**14**(9):e038872.
29. BOLAT I., BITEKER M. *Modified Glasgow Prognostic Score is a novel predictor of clinical outcome in heart failure with preserved ejection fraction*. Scand. Cardiovasc. J. 2020;**54**(3):174–178.
30. CZAPLA M., KWAŚNY A., UCHMANOWICZ I., PIETRZYKOWSKI Ł., LEE C., KOSOWSKI W., et al. *Sex differences in the association between nutritional status and in-hospital mortality in HFpEF patients*. ESC Hear. Fail. 2025;**12**(4):2499–2506.
31. SCHRUTKA L., SEIRER B., RETTL R., DACHS T., BINDER C., DUCA F., et al. *Heart failure with preserved ejection fraction: Calculating the risk of future heart failure events and death*. Front Cardiovasc Med. 2022;**9**:921132.
32. IWAKURA K., ONISHI T., OKAMURA A., KOYAMA Y., TANAKA N., OKADA M., et al. *The WATCH – DM risk score estimates clinical outcomes in type 2 diabetic patients with heart failure with preserved ejection fraction*. Sci Rep. 2024;**14**(1):1746.
33. YEE D., NOVAK E., PLATTS A., NASSIF M.E., LARUE S.J., VADER J.M. *Comparison of the Kansas City Cardiomyopathy Questionnaire and Minnesota Living With Heart Failure Questionnaire in Predicting Heart Failure Outcomes*. Am J Cardiol. 2019;**123**(5):807–812.

Supplementary Table 1

HFpEF definitions used across included studies

Study	HFpEF definition
Rich, 2018 [9]	Framingham HF criteria + LVEF >50%.
Suzuki, 2018 [10]	LVEF ≥50% (HFpEF subgroup).
Thorvaldsen, 2017 [11]	Acute HFpEF hospitalization; LVEF ≥50%.
Matsushita, 2025 [12]	Hospitalized HFpEF; LVEF ≥50%.
Takahari, 2019 [13]	Suspected/exertional HFpEF; preserved EF (LVEF <50% excluded). ^a
Amanai, 2022 [14]	Symptoms + LVEF >50% + objective elevated filling pressures (exercise E/e' or PCWP).
Barros, 2025 [15]	HFpEF defined per authors' diagnostic framework (LVEF preserved). ^a
Hwang, 2020 [16]	Hospitalized HFpEF; preserved EF (LVEF <50% excluded).
Przewłocka-Kośmala, 2022 [17]	Dyspnea + LVEF ≥50% + diastolic dysfunction (sinus rhythm).
Coiro, 2023 [18]	Guideline-defined HFpEF; LVEF ≥50%.
Palazzuoli, 2018 [19]	LVEF ≥50%.
Liang, 2021 [20]	TOPCAT criteria: LVEF ≥45% + HF hospitalization or elevated natriuretic peptides.
Chien, 2019 [21]	Acute HF cohort; preserved EF (LVEF <50% excluded).
Nishi, 2019 [22]	Hospitalized HFpEF; LVEF ≥50%.
Chen, 2023 [23]	HFpEF cohort; preserved EF (LVEF <50% excluded). ^a
Minamisawa, 2019 [24]	TOPCAT criteria: LVEF ≥45% + HF hospitalization or elevated natriuretic peptides.
Zhang, 2025 [25]	ESC 2021 HFpEF criteria; additionally H2FPEF >5.
Kitao, 2025 [26]	Framingham HF + LVEF ≥50% + BNP/NT-proBNP threshold.
Sunaga, 2022 [27]	Framingham HF + LVEF ≥50% + BNP/NT-proBNP threshold.
Lu, 2025 [28]	PARAGON-HF criteria: LVEF ≥45% + elevated natriuretic peptides + structural criteria.
Bolat, 2020 [29]	ESC-based: symptoms/signs + LVEF ≥50% + NP + ≥1 echo criterion.
Czapla, 2025 [30]	HFpEF diagnosis per ESC 2016 guideline (criteria not restated). ^a
Schrutka, 2022 [31]	Symptoms/signs + LVEF ≥50% + diastolic dysfunction + NT-proBNP threshold.
Iwakura, 2024 [32]	Acute HFpEF hospitalization; LVEF ≥50%.
Yee, 2019 [33]	EF ≥50% (HFpEF) vs EF ≤40% (HFrEF comparator).

Note: ^a HFpEF definition not fully restated in the article text; summarized from the authors' stated diagnostic framework, guideline reference, or inclusion/exclusion criteria.

Abbreviations: HFpEF – heart failure with preserved ejection fraction; LVEF – left ventricular ejection fraction; NP – natriuretic peptides; BNP – B-type natriuretic peptide; NT-proBNP – N-terminal pro-B-type natriuretic peptide; PCWP – pulmonary capillary wedge pressure; NR – not reported in the original article.

Supplementary Table 2

Mortality outcomes: prognostic performance of clinician-usable scores in HFpEF cohorts

Study	Score	Setting / cohort	Study type	Population (HFpEF vs mixed-EF)	Mortality endpoint	Follow-up	Discrimination	Effect estimate (unit)	Keynote
Acute admissions: in-hospital and early mortality									
Czapla 2025 [30]	NRS-2002	Hospitalized HFpEF (in-hospital)	Observational screening study	Hospitalized HFpEF	In-hospital mortality	Index hospitalization	NR	Not quantified (model-based odds/probabilities reported; no conventional OR/HR for NRS-2002 provided).	Sex-stratified signal; screening tool rather than HFpEF-specific score
Matsushita 2025 [12]	TRI	Acute HFpEF registry (in-hospital)	Development / validation in registry	Acute HFpEF registry	In-hospital mortality	Index hospitalization	AUC: 0.66 (95% CI, 0.65–0.68).	HR 1.021 per TRI unit (95% CI, 1.014–1.028).	Ultra-low input burden; intended for rapid triage
Thorvaldsen 2017 [11]	ARIC AD-HFpEF	Acute HFpEF admissions (community surveillance)	Derivation + validation (acute setting)	HFpEF admissions (community surveillance)	All-cause mortality	28-day and 1-year	AUC derivation: 0.76 (28-day); 0.72 (1-year). Validation: 0.73 (28-day); 0.71 (1-year).	NR	HFpEF-oriented derivation for acute setting
Hospitalized or post-discharge cohorts: medium-to-long-term mortality									
Hwang 2020 [16]	H2FPEF	Hospitalized HFpEF (long-term follow-up)	Prognostic validation	Hospitalized HFpEF	All-cause mortality	5 years	NR	Adjusted HR 1.053 per 1-point H2FPEF increase (95% CI, 1.008–1.101).	Prognostic gradient may reflect severity phenotype captured by the diagnostic score
Iwakura 2024 [32]	WATCH-DM	Post-discharge HFpEF with T2DM	Phenotype-specific validation	T2DM + HFpEF after hospitalization	All-cause death	Median 386 days	AUC: 0.64.	Adjusted HR 1.102 per 1-point WATCH-DM increase (95% CI, 1.009–1.203).	Diabetes-focused discharge tool; comparator performance similar to general scores
Suzuki 2018 [10]	GWTC-HF	Post-discharge chronic HF (mixed-EF; includes HFpEF)	Validation in post-discharge cohort	Chronic HF after hospitalization (includes HFpEF)	All-cause and cardiac death	Mean 965.8 days	C-stat (95% CI): 0.687 (0.649–0.725) for all-cause death; improved to 0.772 (0.739–0.805) with BNP.	HFpEF subgroup: HR 1.568 per 10-point GWTG-HF increase for all-cause death (95% CI, 1.207–2.023).	Mixed-EF derivation; post-discharge prognostic value in HFpEF subsets

Supplementary Table 2 (continued)

Study	Score	Setting / cohort	Study type	Population (HFpEF vs mixed-EF)	Mortality endpoint	Follow-up	Discrimination	Effect estimate (unit)	Keynote
Trial cohorts									
Lu 2025 [28]	CONUT; GNRI	Trial cohort (PARAGON-HF)	Post-hoc trial analysis (trajectory)	PARAGON-HF HFpEF	CV death	Median 2.9 years	NR	Adjusted HR 1.97 for all-cause death with any abnormal CONUT and/or GNRI (95% CI, 1.65–2.36).	Trajectory-based risk refinement; post-discharge change adds information
Minamisawa 2019 [24]	GNRI	Trial cohort (TOPCAT)	Post-hoc trial analysis	TOPCAT-Americas HFpEF	All-cause death; CV outcomes	Median 2.9 years	NR	HR 1.34–2.06 (by outcome)	Large trial cohort; supports GNRI as baseline risk layer
Diagnostic evaluation cohorts									
Przewlocka-Kośmala 2022 [17]	H2FPPEF	Suspected HFpEF diagnostic cohort (exercise testing)	Head-to-head comparison	Suspected HFpEF with exertional dyspnea (sinus rhythm only)	CV death	Median 48 months	Harrell's C: 0.644	NR	Exercise (step-3) data improved stratification for HFA-PEFF
Przewlocka-Kośmala 2022 [17]	HFA-PEFF (Step 2)	Suspected HFpEF diagnostic cohort (exercise testing)	Head-to-head comparison	Suspected HFpEF with exertional dyspnea (sinus rhythm only)	CV death	Median 48 months	Harrell's C: 0.638	NR	Exercise (step-3) data improved stratification for HFA-PEFF
Przewlocka-Kośmala 2022 [17]	HFA-PEFF (Step 3)	Suspected HFpEF diagnostic cohort (exercise testing)	Head-to-head comparison	Suspected HFpEF with exertional dyspnea (sinus rhythm only)	CV death	Median 48 months	Harrell's C: 0.715	NR	Exercise (step-3) data improved stratification for HFA-PEFF
Przewlocka-Kośmala 2022 [17]	MAGGIC	Suspected HFpEF diagnostic cohort (exercise testing)	Head-to-head comparison	Suspected HFpEF with exertional dyspnea (sinus rhythm only)	CV death	Median 48 months	Harrell's C: 0.637	NR	Exercise (step-3) data improved stratification for HFA-PEFF

Supplementary Table 2 (continued)

Study	Score	Setting / cohort	Study type	Population (HFpEF vs mixed-EF)	Mortality endpoint	Follow-up	Discrimination	Effect estimate (unit)	Keynote
Ambulatory and mixed inpatient/outpatient cohorts									
Bolat 2020 [29]	mGPS (± NT-proBNP)	Ambulatory HFpEF cohort	Observational prognostic study	Ambulatory HFpEF	All-cause death	12 months	AUC 0.759; 0.822 with NT-proBNP	OR 2.42 (mGPS 1) and 3.84 (mGPS 2) vs mGPS 0	Inflammation plus nutrition may complement natriuretic peptides
Coiro 2023 [18]	MEDIA	Hospitalized and ambulatory HFpEF cohorts	Validation across cohorts	Hospitalized HFpEF and stable outpatient HFpEF	All-cause death	Up to 2 years (hospitalized cohort); shorter in ambulatory cohort	Not quantified (qualitative statement only: reported as moderate discrimination; no AUC/C-index provided)	Adjusted HR 2.10 (high vs low score; 95% CI NR)	AF excluded in ambulatory cohort for diastolic validity; limits generalizability
Mixed-EF cohorts: HFpEF subset analyses									
Rich 2018 [9]	MAGGIC (± BNP); SHFM (comparison)	Chronic HF cohort (mixed-EF); HFpEF subset	External validation / comparison	HFpEF subset validation within broader HF cohorts	All-cause mortality	Mean 3.6 years	C-statistic for all-cause mortality: 0.74 (95% CI 0.68–0.80)	NR	Mixed-EF derivation; HFpEF use relies on validation rather than HFpEF-only derivation
Yee 2019 [33]	KCCQ	Chronic HF cohort (mixed-EF; includes HFpEF subset)	Validation / prognostic association	Mixed HF (includes HFpEF subset)	Death/LVAD/transplant	Up to 2 years	C-index: 0.702 (95% CI, 0.666–0.738)	Adjusted HR 0.894 per 5-point KCCQ increase (95% CI, 0.864–0.924)	Patient-reported health status predicts outcomes and may aid shared decisions
Yee 2019 [33]	MLHFQ	Chronic HF cohort (mixed-EF; includes HFpEF subset)	Validation / prognostic association	Mixed HF (includes HFpEF subset)	Death/LVAD/transplant	Up to 2 years	C-index: 0.658 (95% CI, 0.621–0.695)	Adjusted HR 1.077 per 5-point MLHFQ increase (95% CI, 1.045–1.109)	Patient-reported health status predicts outcomes and may aid shared decisions

Supplementary Table 2 (continued)

Study	Score	Setting / cohort	Study type	Population (HFpEF vs mixed-EF)	Mortality endpoint	Follow-up	Discrimination	Effect estimate (unit)	Keynote
Other HFpEF cohorts									
Chen 2023 [23]	CONUT; GNRI; PNI	Older adults with HFpEF	Comparative prognostic study	Older adults with HFpEF	All-cause death	1 year	AUC (all-cause mortality): 0.789 (95% CI, 0.739–0.834).	Cox regression (continuous): CONUT score HR 1.555 (95% CI, 1.419–1.705); GNRI HR 0.949 (95% CI, 0.927–0.971); PNI HR 0.862 (95% CI, 0.824–0.902).	Comparative nutrition indices; mortality associations clearer than readmission
Nishi 2019 [22]	GNRI	HFpEF cohort (setting not specified)	Observational prognostic study	HFpEF	All-cause death	Median 503.5 days	AUROC: GNRI 0.75; albumin 0.71; GNRI+albumin 0.80.	Adjusted HR 3.075 for high GNRI risk (95% CI, 1.244–7.600).	Nutrition-frailty signal prominent in older HFpEF

Abbreviations: AUC – area under the receiver operating characteristic curve; CI – confidence interval; HR – hazard ratio; OR – odds ratio; NR – not reported in the original article; Not quantified – qualitative statement without a numeric estimate; Not extractable – estimates reported only by categories/trajectories without a single summary effect.

Note: Direct numerical comparisons across studies are limited by differences in cohort setting, HFpEF definitions, endpoint definitions, time horizons, and validation design.

Supplementary Table 3

Heart-failure hospitalization and composite outcomes- prognostic performance of clinician-usable scores in HFpEF cohorts

Study	Score	Setting / cohort	Study type	Population	Hospitalization endpoint	Follow-up	Discrimination	Effect estimate (unit)	Keynote
Trial cohorts and stable ambulatory risk									
Coiro 2023 [18]	MEDIA	Hospitalized and ambulatory HFpEF cohorts	Validation across cohorts	Hospitalized and ambulatory HFpEF	HF hospitalization (ambulatory) and composites	Variable by cohort	NR	Adjusted HR 3.39 for HF hospitalization (ambulatory cohort; 95% CI NR).	Echo-only tool; AF excluded in ambulatory cohort
Liang 2021 [20]	C2HEST	Trial cohort (TOPCAT)	Post-hoc trial analysis	TOPCAT HFpEF	HF hospitalization; all-cause hospitalization	Median 3.07 years	Time-dependent AUC (5y): HF hospitalization 0.621 (95% CI, 0.577–0.665); all-cause hospitalization 0.638 (95% CI, 0.606–0.670).	Adjusted HR per 1-point C2HEST increase: HF hospitalization 1.14 (95% CI, 1.01–1.29); any hospitalization 1.22 (95% CI, 1.14–1.29).	Originally AF-focused; here used for broader risk enrichment
Lu 2025 [28]	CONUT; GNRI (baseline and change)	Trial cohort (PARAGON-HF)	Post-hoc trial analysis (trajectory)	PARAGON-HF HFpEF	HF hospitalization	Median 2.9 years	NR	Not extractable (trajectory-specific HRs reported; no single summary effect estimate)	Trajectory-based approach aligns with longitudinal risk refinement
Rich 2018 [9]	MAGGIC ($\pm$ BNP)	Chronic HF cohort (mixed-EF); HFpEF subset	External validation	HFpEF validation subset	Composite: HF hospitalization or death	Mean 3.6 years	C-stat (95% CI): 0.64 (0.58–0.69)	HR 1.8 per 1-SD increase in MAGGIC for composite CV hospitalization or death (95% CI, 1.6–2.1).	Mixed-EF derivation; composite endpoint definitions vary across cohorts
Diagnostic evaluation and prospective HFpEF cohorts									
Barros 2025 [15]	H2FPEF	Prospective HFpEF cohort	Prospective comparison / validation	Prospective HFpEF cohort	HF hospitalization; composite events	2–3 years	AUC: 0.637 (95% CI, 0.518–0.756)	HR 2.316 for H2FPEF >3 vs ≤ 3 (95% CI, 0.973–5.513); $p=0.058$.	Stress testing limited; feasibility influences comparative performance

Supplementary Table 3 (continued)

Study	Score	Setting / cohort	Study type	Population	Hospitalization endpoint	Follow-up	Discrimination	Effect estimate (unit)	Keynote
Barros 2025 [15]	HFA-PEFF	Prospective HFpEF cohort	Prospective comparison / validation	Prospective HFpEF cohort	HF hospitalization; composite events	2–3 years	AUC: 0.572 (95% CI, 0.448–0.696)	NR	Stress testing limited; feasibility influences comparative performance
Przewlocka-Kośmala 2022 [17]	MAGGIC; H2FPPEF; HFA-PEFF	Suspected HFpEF diagnostic cohort (exercise testing)	Head-to-head comparison	Suspected HFpEF with exertional dyspnea	HF hospitalization	Median 48 months	c-index range 0.637–0.715	NR	Exercise testing (step-3) improved prognostic stratification
Hospitalized or post-discharge cohorts: readmission and composite events									
Chen 2023 [23]	CONUT; GNRI; PNI	Older adults with HFpEF	Observational prognostic study	Older HFpEF	HF readmission	1 year	NR	Non-significant association ($p \geq 0.05$; effect estimate not provided)	Readmission links may be weaker than mortality associations in some cohorts
Chien 2019 [21]	GNRI; CONUT; PNI; albumin	Acute HFpEF	Head-to-head comparison	Acute HFpEF	HF rehospitalization	Median 1255 days	NR	Not extractable (index-specific HRs reported; no single summary effect estimate)	Nutrition indices compared head-to-head in acute HFpEF
Hwang 2020 [16]	H2FPPEF	Hospitalized HFpEF (long-term follow-up)	Prognostic validation	Hospitalized HFpEF	Composite: HF hospitalization or all-cause death	5 years	NR	HR 1.05 per 1-point increase	Diagnostic score may capture severity of presentation
Kitao 2025 [26]	CONUT (trajectory)	Acute HFpEF registry (PURSUIT-HFpEF)	Trajectory analysis (registry)	Acute HFpEF (PURSUIT-HFpEF)	HF readmission and death	Mean 482 days	NR	Abnormal CONUT at 1 year: adjusted HR 2.87 (95% CI, 1.32–6.22). CONUT normalization: adjusted HR 0.40 (95% CI, 0.23–0.67).	Serial change adds information beyond baseline score

Supplementary Table 3 (continued)

Study	Score	Setting / cohort	Study type	Population	Hospitalization endpoint	Follow-up	Discrimination	Effect estimate (unit)	Keynote
Palazzuoli 2018 [19]	LUS B-lines ($\pm$ BNP, echo)	Acute HF (mixed-EF) at discharge	Observational prognostic study	Acute HF (HFpEF and HFrEF)	Composite: 6-month readmission or death	6 months	AUC (death/HF rehospitalization): 0.83 (95% CI, 0.77–0.90).	Multivariable HR 1.16 per additional B-line at discharge (95% CI, 1.11–1.21).	Discharge congestion may outperform admission-only measures Youden cut-point: 22 B-lines at discharge.
Sunaga 2022 [27]	GNRI; CONUT; PNI (change during stay)	Hospitalized HFpEF	Change-based prognostic analysis	Hospitalized HFpEF	HF hospitalization and death	Median 421 days	NR	Not quantified (directionally consistent associations; no single summary numeric estimate)	Emphasizes in-hospital change rather than baseline alone

Abbreviations: AUC – area under the receiver operating characteristic curve; CI – confidence interval; HFH – heart failure hospitalization; HR – hazard ratio; IR – incidence rate; NR – not reported in the original article; Not quantified – qualitative statement without a numeric estimate; Not extractable – estimates reported only by categories/trajectories without a single summary effect.

Note: Direct numerical comparisons across studies are limited by differences in cohort setting, HFpEF definitions, endpoint definitions, time horizons, and validation design.