Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

From risk marker to treatment trigger: biomarker-guided decision-making in acute heart failure

Mykhailo Sosnov

msosnov97@gmail.com

Bogomolets National Medical University

Mariia Seleznova

Bogomolets National Medical University

Systematic Review

Keywords: acute heart failure, biomarkers, natriuretic peptides, decongestion, iron deficiency, risk stratification

Posted Date: August 26th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10629275/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

Acute heart failure (AHF) biomarkers are overwhelmingly validated as predictors of outcome, yet the literature rarely distinguishes prognostic value from a genuinely different question: whether a biomarker result should change a specific clinical decision. This narrative review, based on a PubMed/MEDLINE search of the English-language literature from January 2019 to July 2026, examines the evidence for biomarkers as treatment triggers rather than risk labels across several domains. Natriuretic peptides remain the best-established diagnostic aid in AHF, but trials using them to guide chronic outpatient therapy intensification have produced inconsistent results, and their accuracy is reduced by obesity, atrial fibrillation, and renal dysfunction. Congestion biomarkers, including carbohydrate antigen 125 (CA-125), adrenomedullin and its derivatives, and soluble suppression of tumorigenicity-2, correlate consistently with residual congestion and prognosis; a CA-125-guided decongestion strategy has already shown a significant reduction in death or readmission across two single-center randomized trials from the same investigator group, although this finding has not yet been replicated independently or adopted into guidelines. Iron-deficiency biomarkers are the most clinically mature example of a guideline-embedded treatment trigger, yet across seven major trials spanning 2009 to 2025, results have been inconsistent, funding has come predominantly from the manufacturers of the therapies being tested, and a pre-specified sex-differential effect in the largest recent trial raises the question of whether a single ferritin/transferrin-saturation threshold should apply uniformly. Newer multi-marker panels add prognostic precision without an accompanying management action. The clearest example of a biomarker functioning as an actual decision trigger is the STRONG-HF trial, in which serial N-terminal pro-B-type natriuretic peptide measurements were incorporated directly into eligibility criteria and safety rules governing titration of guideline-directed medical therapy. This review argues that future biomarker research in AHF should be explicitly designed around decision impact, and proposes that congestion biomarkers, building on the CHANCE-HF proof of concept, are the most immediately tractable candidates for a larger, STRONG-HF-style embedded trial.

Background

Heart failure (HF) is a clinical syndrome defined by signs and symptoms caused by a structural or functional cardiac abnormality, corroborated by elevated natriuretic peptide levels or objective evidence of cardiogenic congestion [1]. Acute heart failure (AHF) — a rapid onset or worsening of these signs and symptoms requiring urgent evaluation and treatment — remains one of the most common reasons for hospital admission in cardiovascular medicine, and carries a persistently high risk of readmission and death despite decades of trial activity [2, 3].

Over the same decades, the number of circulating biomarkers reported to associate with AHF diagnosis, severity, or outcome has grown very large: natriuretic peptides, cardiac troponins, soluble suppression of tumorigenicity-2 (sST2), growth differentiation factor-15 (GDF-15), carbohydrate antigen 125 (CA-125), adrenomedullin (ADM) and its derivatives, soluble urokinase plasminogen activator receptor (suPAR), fibroblast growth factor-23 (FGF-23), and many others each have a published evidence base linking them

to prognosis [4, 5]. This literature is valuable, but it is also strikingly homogeneous in its framing: the recurring question is whether a marker correlates with an outcome, not whether measuring it should change what a clinician actually does.

These are different questions with different evidentiary requirements. A biomarker can be robustly prognostic — reliably separating higher-risk from lower-risk patients — without there being any evidence that acting on its result (adjusting a diuretic dose, initiating intravenous iron, accelerating guideline-directed medical therapy [GDMT] titration, timing hospital discharge) improves outcomes. Demonstrating the latter requires a different kind of study: one in which the biomarker itself is embedded in the intervention being tested, either as an eligibility criterion, a titration trigger, or a safety stopping rule, and is compared against a strategy that does not use it.

Search strategy and scope

This narrative review is based on a PubMed/MEDLINE search covering the English-language literature published between January 2019 and July 2026, combining “acute heart failure” with each of “biomarker”, “natriuretic peptide”, “congestion”, “iron deficiency”, “CA-125”, “adrenomedullin”, “soluble ST2”, “GDF-15”, “suPAR”, and “guideline-directed medical therapy”. Reference lists of identified systematic reviews and major trial publications were hand-searched for additional primary sources, and landmark trials published before 2019 (e.g. the original FAIR-HF and CHANCE-HF trials) were retained where they established a definition, threshold, or proof-of-concept still relevant to current practice. The review does not attempt a systematic or exhaustive count of every AHF biomarker study; it is deliberately selective, prioritizing trials and systematic reviews that speak directly to whether a biomarker result is, or could be, linked to a specific clinical action, over the much larger body of purely observational prognostic association studies. Generative AI (Claude, Anthropic) was used to assist with literature searching, drafting, and verification of statistical and bibliographic details during preparation of this manuscript. All scientific content, interpretations, and conclusions were reviewed and approved by the authors, who take full responsibility for the accuracy and integrity of the work.

The aim of this review is therefore to examine, across the major biomarker domains studied in AHF, how much of the existing evidence actually supports the stronger, decision-relevant claim, as opposed to the weaker and much more common claim of prognostic association. We review natriuretic peptides, congestion biomarkers, iron-deficiency biomarkers, and next-generation multi-marker panels in turn, examine the one AHF trial programmed — STRONG-HF — in which a biomarker was designed into the treatment algorithm itself, and close by discussing the barriers, including trial funding and assay standardization, that currently limit translation of the others from risk marker to treatment trigger.

To avoid an impressionistic “ready/not ready” judgement, Table 1 classifies clinical readiness against an explicit four-level rubric based purely on trial design: (0) prognostic association only, with no prospective interventional evidence of any kind; (1) at least one prospective interventional signal, typically a single-center or investigator-led trial, but no guideline adoption; (2) a biomarker checked once, at trial entry, as a

static eligibility gate, with no further biomarker-guided decision made afterwards; (3) a biomarker measured repeatedly during the trial and used to drive an ongoing decision — randomization, titration, or a safety rule — compared against a strategy without it. The distinction between (2) and (3) is not whether a biomarker threshold appears in a trial at all, but whether it continues to act on the patient's management after enrolment: a one-off entry criterion is scored 2 however central it is to the trial's design, while a marker re-measured to guide what happens next is scored 3. This score is deliberately a measure of design tier alone; it does not by itself capture whether that design has been independently replicated. Where a level-3 design exists but originates from a single funding source or investigator group without independent confirmation, this is flagged separately with an asterisk and explained in the table footnote, rather than folded into a lower numeric score, since conflating the two would obscure exactly the distinction this review is trying to draw out. The rubric is applied identically to every biomarker in Table 1, including those discussed favorably elsewhere in this review; the ferritin/transferrin-saturation threshold, for instance, functions only as a one-off entry gate across all seven iron trials in Table 2 and is scored accordingly, despite being central to those trials' designs.

Table 1
Acute heart failure biomarkers: prognostic evidence versus decision-impact evidence

Biomarker(s)	**Prognostic evidence**	**Decision-impact evidence**	**Clinical readiness (0–3)**
NT-proBNP / BNP	Extensive; validated diagnostic thresholds [6, 7]	Mixed for outpatient titration [10]; embedded as eligibility/safety trigger in STRONG-HF [28–30]	2 (diagnosis) / 3* (STRONG-HF)
CA-125	Correlates with serosal congestion, mortality, readmission [11, 12]	Positive in 2 single-centre RCTs from one investigator group; not independently replicated [15, 16]	3* (single-source)
Bioactive adrenomedullin	Reflects rapid haemodynamic change; identifies residual congestion [13]	No dedicated decision trial	0
sST2	Independent prognostic value; tracks decongestion [12, 14]	No dedicated decision trial	0
GDF-15 / suPAR / FGF-23	Each independently associated with mortality [4]	None identified	0
Ferritin + transferrin saturation	Defines iron-deficiency phenotype linked to symptoms and prognosis [17]	Explicit RCT eligibility criterion across 7 trials [18–22, 24, 26]; sex-differential effect recently reported [25]	2 (under re-appraisal)

**Design tier 3 achieved, but from a single funding source and/or investigator group without independent replication (see main text). BNP, B-type natriuretic peptide; CA-125, carbohydrate antigen 125; FGF-23, fibroblast growth factor-23; GDF-15, growth differentiation factor-15; NT-proBNP, N-terminal pro-B-type natriuretic peptide; RCT, randomised controlled trial; sST2, soluble suppression of tumorigenicity-2; suPAR, soluble urokinase plasminogen activator receptor.*

Natriuretic peptides: established diagnostic role, unresolved decision-making role

B-type natriuretic peptide (BNP) and N-terminal pro-B-type natriuretic peptide (NT-proBNP) are the best-validated biomarkers in AHF. International consensus guidance from the Heart Failure Association of the European Society of Cardiology sets out eleven key principles for their use, foremost among them that natriuretic peptides should be measured in all patients presenting with symptoms suggestive of HF, always alongside other clinical information rather than in isolation [6]. Their diagnostic performance is well characterized: at thresholds of approximately 100 pg/mL for BNP and 300 pg/mL for NT-proBNP, sensitivity and negative predictive value for AHF both exceed 90%, making a low result useful for ruling the diagnosis out [7]. Their accuracy, however, is measurably reduced by obesity, atrial fibrillation, and renal dysfunction, all of which are common in the AHF population and complicate interpretation of borderline values [7].

Where the evidence is markedly less settled is in the use of natriuretic peptides not merely to diagnose AHF, but to guide the intensity or timing of treatment. The concept of natriuretic peptide-guided management — adjusting chronic HF therapy according to serial measurements rather than symptoms alone — has been reviewed extensively [8], and individual-patient data suggest that a greater therapy-induced fall in natriuretic peptide concentration is associated with better long-term survival [9]. Despite this biological plausibility, the largest outpatient trial of NT-proBNP-guided titration of chronic HF therapy did not demonstrate a reduction in the primary composite of hospitalization or cardiovascular death when therapy was titrated to a natriuretic peptide target compared with usual clinical judgement [10]. This inconsistency matters for the present review's framing: NT-proBNP is unambiguously useful for diagnosis, and its trajectory correlates with prognosis, but the additional step — using that trajectory to trigger a specific treatment action in the outpatient setting — has a much thinner and more mixed evidence base, a distinction addressed directly in the STRONG-HF discussion below, where the same biomarker was used quite differently and much more successfully.

Congestion biomarkers and decongestion decisions

Congestion, not merely elevated filling pressure, is now understood as the central pathophysiological driver of AHF symptoms and a major determinant of readmission [11]. A large registry analysis of the ESC Heart Failure Long-Term Registry has confirmed that the clinical profile of congestion and perfusion at presentation carries independent prognostic weight for both in-hospital and long-term outcomes [3]. Because natriuretic peptides provide only an indirect readout of congestion and are influenced by

numerous non-cardiac factors, several other biomarkers have been studied specifically as measures of congestion itself, with the explicit hope that they might guide decongestive therapy rather than simply predict outcome.

CA-125, better known as a tumor marker, is released by mesothelial cells under mechanical stretch and correlates with serosal (pleural, peritoneal, pericardial) fluid accumulation; admission levels predict mortality and readmission, and its slower kinetics compared with natriuretic peptides may make it a useful marker of more chronic, higher-volume congestion [11, 12]. Bioactive ADM, a vasodilatory peptide with a very short half-life, reflects rapid hemodynamic change and has shown a role in identifying residual, subclinical congestion at hospital discharge, a state associated with early readmission that is easy to miss on clinical examination alone [11, 13]. sST2, released in response to myocardial strain and inflammation, adds prognostic information largely independent of natriuretic peptides and shows a dynamic decline during effective decongestion that has been proposed as a marker of treatment response [11, 14]. A large proteomic analysis of the BIOSTAT-CHF cohort similarly found that severe congestion — defined clinically by jugular venous pressure, orthopnea, and peripheral oedema — was associated with up-regulation of several of these same markers (CA-125, sST2, GDF-15, and others), reinforcing that congestion has a distinct, measurable biological signature beyond the natriuretic peptide axis [5].

Unlike the other congestion biomarkers discussed here, CA-125 has already been tested as an explicit decision trigger rather than only as a correlate of congestion. In the CHANCE-HF trial, 380 patients discharged after AHF with elevated CA-125 were randomized to a CA-125-guided strategy — using serial measurements to titrate diuretic and statin doses and to intensify monitoring — or standard care; the CA-125-guided strategy significantly reduced the composite of one-year death or AHF readmission, an effect that was independent of changes in natriuretic peptides [15]. A related single-center trial in patients with AHF and renal dysfunction similarly found that a CA-125-guided diuretic strategy improved renal function parameters at 72 hours compared with usual care [16].

Two features of this evidence base temper how much weight it can currently bear, and it is important to apply the same scrutiny here that we apply to the iron trials in the following section. First, both CA-125 trials were sponsored by the same academic institution (the Fundación para la Investigación del Hospital Clínico de Valencia) and led by the same principal investigator, with substantially overlapping investigator teams [15, 16]. This is a materially different concern from the industry funding discussed for intravenous iron below, but it is a concern of the same kind: a single, non-independent source of evidence for a proposed decision rule, however methodologically sound, is not equivalent to independent multicenter replication, and neither trial has yet been repeated by a separate group. Second, both trials were open-label: treating physicians and patients were aware of CA-125 values and treatment assignment, and the intervention itself consisted largely of process measures — more frequent visits, more active diuretic and statin titration — that are inherently difficult to blind and are themselves susceptible to performance bias independent of any true biological effect of CA-125-guided care. Neither limitation invalidates the finding, but both mean that CHANCE-HF should be read as a genuine proof of

concept that the STRONG-HF template is achievable outside the GDMT-titration setting, not as definitive, guideline-ready evidence. What is still missing is independent, multicenter, and ideally blinded-endpoint confirmation at a scale sufficient to change practice.

Biomarker-guided iron replacement: a cautionary re-appraisal

Iron deficiency, defined biochemically rather than by hemoglobin, is present in roughly half of patients with HF and has an established association with worse symptoms, exercise capacity, and prognosis independent of anemia [17]. Intravenous ferric carboxymaltose (FCM) is recommended by international guidelines for iron-deficient patients with heart failure with reduced ejection fraction (HFrEF), and this recommendation itself illustrates the central theme of this review: the guideline recommendation follows directly from a biochemical definition of iron deficiency (serum ferritin below 100 ng/mL, or 100–299 ng/mL with transferrin saturation below 20%) that functions as an explicit treatment trigger, not merely a prognostic marker.

This threshold, however, has been tested across seven major randomized trials spanning 16 years with an inconsistent pattern of results (Table 2). The original FAIR-HF trial in 2009 established both the ferritin/transferrin-saturation definition and a positive symptomatic effect [18]. CONFIRM-HF subsequently showed sustained improvement in functional capacity and quality of life with long-term FCM in chronic HFrEF [19]. AFFIRM-AHF, testing FCM given at discharge after AHF, showed a primary composite result (52.5% vs 67.6% for total HF hospitalizations and cardiovascular death) that favored FCM but narrowly missed the pre-specified significance threshold ($p = 0.059$) [20]. IRONMAN, the only investigator-initiated trial in this group rather than an industry-sponsored one, suggested a reduction in recurrent HF hospitalisation with ferric derisomaltose compared with usual care [21]. HEART-FID, the largest trial to date at over 3,000 patients, did not meet its hierarchical primary endpoint [22]. An updated meta-analysis pooling these and other trials found a consistent, if individually modest, direction of benefit for intravenous iron on hospitalization and quality-of-life outcomes [23]. Most recently, FAIR-HF2 again showed a primary composite result favoring FCM (a 21% relative reduction in cardiovascular death or first HF hospitalization) that did not reach statistical significance [24], and a pre-specified sex-specific analysis of the same trial found a more pronounced benefit in men (hazard ratio 0.86) than in women, in whom there was no clear benefit and a non-significant signal towards increased mortality (hazard ratio 1.46) [25]. In HFpEF, FAIR-HFpEF was stopped early after enrolling only 39 of a planned 200 patients, though the primary endpoint (six-minute walk distance) favored FCM by a clinically meaningful margin in this small sample [26, 27].

Two points follow directly from this pattern. First, six of the seven trials in Table 2 were sponsored by the manufacturer or licensed distributor of the ferric carboxymaltose or ferric derisomaltose product under study; only IRONMAN was investigator-initiated. This is not evidence that the trials were conducted improperly, but it is directly relevant to the theme of this review: when the biochemical eligibility threshold used to select patients for an intervention is developed and tested predominantly in trials funded by the manufacturer of that intervention, independent replication and independent re-examination

of the threshold itself — exactly the kind of re-examination that produced the FAIR-HF2 sex-differential finding — take on particular importance. Second, and more directly, the sex-specific finding in FAIR-HF2 shows that an existing, guideline-embedded biomarker threshold may not generalize uniformly even within the population it was designed for, which is a materially different and more troubling problem than a single trial narrowly missing significance.

Table 2
Major intravenous iron trials in heart failure: eligibility criteria, results, and funding source

Trial (year)	N	Population	Primary outcome result	Funding source
FAIR-HF (2009) [18]	459	Chronic HFrEF, iron deficiency	Improved symptoms and 6MWD vs placebo	Vifor Pharma
CONFIRM-HF (2015) [19]	304	Chronic HFrEF, iron deficiency	Improved 6MWD and quality of life vs placebo	Vifor Pharma
AFFIRM-AHF (2020) [20]	1,132	Discharge after AHF, iron deficiency	52.5% vs 67.6% HF hospitalisation/CV death; p = 0.059 (not significant)	Vifor Pharma
IRONMAN (2022) [21]	1,137	Chronic HFrEF, iron deficiency	Reduced recurrent HF hospitalisation vs usual care	Investigator-initiated (UK NIHR)
HEART-FID (2023) [22]	3,065	Chronic HFrEF, iron deficiency	Hierarchical composite not met	American Regent (Daiichi Sankyo)
FAIR-HF2 (2025) [24, 25]	1,105	Chronic HFrEF, iron deficiency	21% relative reduction, not significant; benefit in men only	German Centre for Cardiovascular Research (DZHK), academic
FAIR-HFpEF (2024) [26, 27]	39 of planned 200	Chronic HFpEF, iron deficiency	6MWD improved; stopped early for slow recruitment	Vifor Pharma

6MWD, six-minute walk distance; AHF, acute heart failure; CV, cardiovascular; HFpEF, heart failure with preserved ejection fraction; HFrEF, heart failure with reduced ejection fraction; NIHR, National Institute for Health and Care Research.

Next-generation multi-marker panels: prognosis versus action

A growing panel of newer biomarkers — GDF-15, sST2, suPAR, and FGF-23 among them — has been investigated in AHF and HFrEF cohorts, generally in combination with NT-proBNP, and generally with the finding that each adds independent prognostic value for mortality or readmission [4]. A representative recent study followed 104 patients with acute HFrEF decompensation for a median of 23.5 months and found that NT-proBNP, GDF-15, sST2, suPAR, and FGF-23 were each independently associated with all-cause mortality [4].

What this and comparable studies do not show is any accompanying change in management triggered by these markers: none of the panels reviewed here has, to date, been incorporated into a trial in which a specific therapeutic action (a change in diuretic strategy, a referral decision, a threshold for advanced therapies) was contingent on the marker result and tested against usual care. This is not a criticism of the underlying biology — the mechanistic case for markers of inflammation (GDF-15, suPAR) and mineral metabolism (FGF-23) contributing independently to risk is reasonably strong — but it is a direct illustration of the gap this review is concerned with: an expanding evidence base for prognosis, with almost no corresponding evidence base for action, in clear contrast to the CA-125 trials described above.

STRONG-HF: a biomarker written into the treatment algorithm

Registry data have repeatedly shown that patients hospitalized with HF are frequently discharged without full doses of GDMT, and that outpatient up-titration afterwards is slow and often incomplete [3]. The STRONG-HF trial directly tested whether this could be addressed by a structured, biomarker-informed strategy of rapid in-hospital and early post-discharge up-titration, with its rationale and design published separately from the main results [28, 29]. Patients not on full-dose GDMT at admission were randomized to a high-intensity care strategy — with GDMT up-titrated to full doses within two weeks of discharge under close safety monitoring — or usual care; the high-intensity strategy reduced the composite of 180-day all-cause death or HF readmission and the trial was stopped early because the between-group difference was larger than anticipated [28].

Table 3
STRONG-HF: NT-proBNP embedded as an eligibility criterion and safety trigger, not a prognostic covariate

Trial stage	NT-proBNP-based rule	Consequence for the patient
Randomisation eligibility	Screening NT-proBNP > 1,500 pg/mL that had fallen by > 10% before randomisation	Selected patients with persistent but improving congestion, excluding those too early or already stable
Up-titration visits	One of five pre-specified safety indicators (with eGFR, potassium, systolic BP, heart rate)	A > 10% rise above pre-discharge value triggered a mandatory delay in further GDMT dose escalation
Observed impact	> 50% of high-intensity arm met ≥ 1 safety indicator during follow-up [30]	Meeting a safety indicator was associated with a lower achieved GDMT dose, confirming the rule actively shaped treatment

BP, blood pressure; eGFR, estimated glomerular filtration rate; GDMT, guideline-directed medical therapy; NT-proBNP, N-terminal pro-B-type natriuretic peptide.

This is, in the terms set out in this review, the clearest existing example of a biomarker functioning as a genuine treatment trigger in AHF: NT-proBNP determined who could be randomized, and its trajectory during follow-up determined whether GDMT titration could proceed or had to be paused. A pre-specified

sub-analysis restricted to elderly patients found that the same NT-proBNP-guided high-intensity strategy remained safe and effective in this higher-risk group, supporting generalizability of the approach across age [31].

Applying the same funding scrutiny here that we apply to the iron trials in the next section: “investigator-initiated” describes who designed and controlled the trial (an academic steering committee, here operating through the non-profit Heart Initiative), not who paid for it, and the two are not mutually exclusive. STRONG-HF was an investigator-initiated study, but it was funded through a grant from, and supported with diagnostic products by, Roche Diagnostics — the manufacturer of the Elecsys NT-proBNP assay used to apply the trial's eligibility and safety rules — and the trial's lead author has disclosed research grants and lecture honoraria from Roche Diagnostics among other companies [28]. This does not on its own cast doubt on the result, which was large, pre-specified, and led to early termination for efficacy, but it means STRONG-HF sits under the same structural caveat as the iron-deficiency threshold discussed below: the clearest example in this review of a biomarker embedded in a treatment algorithm was also developed and funded in partnership with the diagnostics manufacturer whose assay defines that algorithm, and independent confirmation using a different NT-proBNP platform, or a different manufacturer's assay entirely, would strengthen confidence that the finding reflects the biomarker strategy rather than any feature specific to that funding relationship.

The lesson for the other biomarker domains reviewed above is structural as much as clinical — it is the explicit incorporation of the biomarker into the intervention and its comparison against a strategy without it, not the marker's prognostic strength per se, that generates decision-relevant evidence. Two recently approved therapies for HF with preserved or mildly reduced ejection fraction — the non-steroidal mineralocorticoid receptor antagonist finerenone and the incretin-based agent tirzepatide — have not yet been tested within any comparable biomarker-embedded titration framework, and their addition to an already multi-drug regimen makes the STRONG-HF question more, not less, urgent to answer for congestion biomarkers specifically [32–35].

Barriers to clinical translation

Several structural barriers explain why the decision-focused evidence base remains so much thinner than the prognostic literature. First, generating it requires a fundamentally different trial design: rather than measuring a biomarker and following patients to an outcome, the biomarker itself must be written into the protocol as an eligibility criterion or a titration or safety rule, and compared head-to-head against a strategy that omits it. STRONG-HF and the CHANCE-HF trials are templates for this design; almost none of the remaining congestion or multi-marker literature reviewed above has been tested this way, and even the positive CA-125 trials remain unreplicated outside a small number of overlapping investigator groups.

Second, and as Table 2 makes explicit, funding structure is itself a barrier to independent threshold validation: six of the seven major intravenous iron trials were sponsored by the manufacturer or licensed

distributor of the product under investigation, and the one exception, IRONMAN, was also the trial that most directly prompted debate about whether the ferritin/transferrin-saturation definition travels well across settings. This is not a problem unique to iron: STRONG-HF, this review's central example of a successful biomarker-embedded trial, was funded by the manufacturer of the assay embedded in its protocol, and both CA-125 decision trials originated from a single academic group without independent replication. Consistency requires acknowledging that none of the three most decision-relevant findings in this review — STRONG-HF, CHANCE-HF, and the iron-deficiency threshold — currently has fully independent, multi-source confirmation. Independent, differently funded replication should be treated as a priority for all three, not applied selectively to the domain that happens to have the least favorable findings.

Third, assay heterogeneity across centers and platforms complicates any attempt to standardize a decision threshold, particularly for newer markers such as CA-125-based congestion assessment or sST2, where cut-points have generally been derived retrospectively within individual cohorts rather than prospectively validated as decision boundaries [11, 12]. A concrete regulatory precedent for how this could be addressed already exists, though it requires distinguishing two separate FDA programmes rather than treating them as one: the Center for Drug Evaluation and Research's Biomarker Qualification Program, established under the 21st Century Cures Act, provides a defined three-stage process (Letter of Intent, Qualification Plan, Full Qualification Package) for qualifying a biomarker for a specific context of use in drug development, while a structurally similar but administratively distinct pathway exists through the Center for Devices and Radiological Health for diagnostic assays; cardiac troponin I's established role in myocardial infarction diagnosis was established through the latter, device-focused route, not the drug-development Biomarker Qualification Program itself [36]. No congestion biomarker discussed in this review has entered this or an equivalent process for a decongestion-guidance context of use; doing so, rather than continuing to rely on cohort-specific retrospective cut-points, would materially shorten the distance between an analytically sound assay and a clinically actionable threshold.

Finally, several of the markers with the most mechanistic promise — GDF-15, suPAR, FGF-23 — are non-specific to the heart and are influenced by renal function, inflammation, and comorbidity burden, which is a further obstacle to defining a single actionable threshold applicable across the heterogeneous AHF population [4]. Cost and accessibility, particularly for assays not yet in widespread clinical use outside specialist centers, remain a more familiar but still substantial barrier to any biomarker-guided strategy reaching routine AHF care regardless of the strength of its evidence base.

Conclusions

Biomarker research in acute heart failure has produced a large and still-growing evidence base for prognosis, but a much smaller one for decision-making, and the two should not be conflated when judging a biomarker's clinical maturity. Natriuretic peptides are excellent diagnostic aids but have an inconsistent record when used to guide chronic therapy titration in isolation. Iron-deficiency biomarkers are the most mature example of a guideline-embedded treatment trigger, yet the pattern across seven

trials in Table 2 — predominantly industry-funded, inconsistently positive, and now showing a sex-differential effect in the largest recent trial — argues for independently funded re-examination of the existing threshold rather than further trials using the same definition. Next-generation multi-marker panels currently add prognostic precision without an accompanying management action.

Of the domains reviewed here, congestion biomarkers are, in our view, the most immediately tractable candidates for a larger STRONG-HF-style embedded trial. Setting NT-proBNP/STRONG-HF aside as the review's central exemplar, congestion biomarkers are the only one of the remaining domains for which a proof of concept already exists: CHANCE-HF and its follow-up renal-dysfunction trial showed that a CA-125-guided decongestion strategy can be embedded directly into randomization and produce a significant reduction in death or readmission. Unlike GDF-15, suPAR, and FGF-23, congestion biomarkers have a specific, testable clinical action attached to them (decongestive therapy itself, rather than a generic GDMT algorithm), and unlike the iron-deficiency threshold, they do not yet carry a fixed guideline-embedded definition that would need to be undone first. A plausible next step would be a pragmatic, multicenter trial that replicates and extends the CHANCE-HF design across a broader range of centers and patient populations, using a congestion biomarker such as CA-125, bioactive ADM, or sST2, measured serially during and after an AHF admission, to trigger diuretic dose adjustment or determine discharge readiness, compared against usual clinical assessment. Patient-centered endpoints such as days alive and out of hospital, or validated quality-of-life instruments, alongside conventional hospitalization and mortality outcomes, would allow such a trial to confirm decision impact at a scale sufficient to inform guidelines. Until trials of this design exist across the remaining biomarker domains reviewed here, claims that a given marker is ready to “guide therapy” in AHF should be treated as aspirational rather than established.

Abbreviations

ADM
adrenomedullin
AHF
acute heart failure
BNP
B-type natriuretic peptide
CA-125
carbohydrate antigen 125
DZHK
German Centre for Cardiovascular Research
FCM
ferric carboxymaltose
FGF-23
fibroblast growth factor-23
GDF-15

growth differentiation factor-15
GDMT
guideline-directed medical therapy
HF
heart failure
HFpEF
heart failure with preserved ejection fraction
HFrEF
heart failure with reduced ejection fraction
NT-proBNP
N-terminal pro-B-type natriuretic peptide
sST2
soluble suppression of tumorigenicity-2
suPAR
soluble urokinase plasminogen activator receptor
6MWD
six-minute walk distance.

Declarations

Author Contribution

Mykhailo Sosnov conceived the review, performed the literature search, and drafted the manuscript. Mariia Seleznova contributed to the literature search and critically revised the manuscript. Both authors read and approved the final manuscript.

Acknowledgement

Not applicable.

Data Availability

Not applicable

References

1. Bozkurt B, Coats AJS, Tsutsui H et al (2021) Universal definition and classification of heart failure: a report of the Heart Failure Society of America, Heart Failure Association of the European Society of Cardiology, Japanese Heart Failure Society and Writing Committee of the Universal Definition of Heart Failure. J Card Fail 27(4):387–413

2. McDonagh TA, Metra M, Adamo M et al (2021) 2021 ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure. Eur Heart J 42(36):3599–3726
3. Chioncel O, Mebazaa A, Harjola VP et al (2017) Clinical phenotypes and outcome of patients hospitalized for acute heart failure: the ESC Heart Failure Long-Term Registry. Eur J Heart Fail 19(10):1242–1254
4. Cortés M, Lumpuy-Castillo J, García-Talavera CS et al (2025) New biomarkers in the prognostic assessment of acute heart failure with reduced ejection fraction: beyond natriuretic peptides. Int J Mol Sci 26(3):986
5. Pandhi P, Ter Maaten JM, Anker SD et al (2022) Pathophysiologic processes and novel biomarkers associated with congestion in heart failure. JACC Heart Fail 10(9):623–632
6. Mueller C, McDonald K, de Boer RA et al (2019) Heart Failure Association of the European Society of Cardiology. Heart Failure Association of the European Society of Cardiology practical guidance on the use of natriuretic peptide concentrations. Eur J Heart Fail 21(6):715–731
7. Richards AM (2017) Biomarkers in acute heart failure - cardiac and kidney. Card Fail Rev 3(2):89–94
8. Chioncel O, Collins SP, Greene SJ et al (2016) Natriuretic peptide-guided management in heart failure. J Cardiovasc Med 17(8):556–568
9. Wessler BS, McCauley M, Morine K, Konstam MA, Udelson JE (2019) Relation between therapy-induced changes in natriuretic peptide levels and long-term therapeutic effects on mortality in patients with heart failure and reduced ejection fraction. Eur J Heart Fail 21(5):613–620
10. Felker GM, Anstrom KJ, Adams KF et al (2017) Effect of natriuretic peptide-guided therapy on hospitalization or cardiovascular mortality in high-risk patients with heart failure and reduced ejection fraction: the GUIDE-IT randomized clinical trial. JAMA 318(8):713–720
11. de la Espriella R, Núñez-Marín G, Codina P, Núñez J, Bayés-Genís A (2023) Biomarkers to improve decision-making in acute heart failure. Card Fail Rev 9:e08
12. Núñez J, de la Espriella R, Rossignol P et al (2022) Congestion in heart failure: a circulating biomarker-based perspective. A review from the Biomarkers Working Group of the Heart Failure Association, European Society of Cardiology. Eur J Heart Fail 24(10):1751–1766
13. Egerstedt A, Czuba T, Bronton K et al (2022) Bioactive adrenomedullin for assessment of venous congestion in heart failure. ESC Heart Fail 9(5):3543–3555
14. Correale M, Tricarico L, Fioretti F, Croella F, Inciardi RM, Mattioli AV (2023) The role of congestion biomarkers in heart failure with reduced ejection fraction. J Clin Med 12(11):3834
15. Núñez J, Llàcer P, Bertomeu-González V et al (2016) Carbohydrate antigen-125-guided therapy in acute heart failure: CHANCE-HF: a randomized study. JACC Heart Fail 4(11):833–843
16. Núñez J, Llàcer P, García-Blas S et al (2020) CA125-guided diuretic treatment versus usual care in patients with acute heart failure and renal dysfunction. Am J Med 133(3):370–380e4
17. von Haehling S, Ebner N, Evertz R, Ponikowski P, Anker SD (2019) Iron deficiency in heart failure: an overview. JACC Heart Fail 7(1):36–46

18. Anker SD, Comin Colet J, Filippatos G et al (2009) FAIR-HF Trial Investigators. Ferric carboxymaltose in patients with heart failure and iron deficiency. N Engl J Med 361(25):2436–2448
19. Ponikowski P, van Veldhuisen DJ, Comin-Colet J et al (2015) CONFIRM-HF Investigators. Beneficial effects of long-term intravenous iron therapy with ferric carboxymaltose in patients with symptomatic heart failure and iron deficiency. Eur Heart J 36(11):657–668
20. Ponikowski P, Kirwan BA, Anker SD et al (2020) AFFIRM-AHF Investigators. Ferric carboxymaltose for iron deficiency at discharge after acute heart failure: a multicentre, double-blind, randomised, controlled trial. Lancet 396(10266):1895–1904
21. Kalra PR, Cleland JGF, Petrie MC et al (2022) Intravenous ferric derisomaltose in patients with heart failure and iron deficiency in the UK (IRONMAN): an investigator-initiated, prospective, randomised, open-label, blinded-endpoint trial. Lancet 400(10369):2199–2209
22. Mentz RJ, Garg J, Rockhold FW et al (2023) HEART-FID Investigators. Ferric carboxymaltose in heart failure with iron deficiency. N Engl J Med 389(11):975–986
23. Graham FJ, Pellicori P, Kalra PR, Ford I, Bruzzese D, Cleland JGF (2023) Intravenous iron in patients with heart failure and iron deficiency: an updated meta-analysis. Eur J Heart Fail 25(4):528–537
24. Anker SD, Friede T, Butler J et al (2025) Intravenous ferric carboxymaltose in heart failure with iron deficiency: the FAIR-HF2 DZHK05 randomized clinical trial. JAMA 333(22):1965–1976
25. Karakas M, Friede T, Butler J et al (2025) Intravenous ferric carboxymaltose in heart failure with iron deficiency (FAIR-HF2 DZHK05 trial): sex-specific outcomes. Eur J Heart Fail. Epub ahead of print.
26. von Haehling S, Doehner W, Evertz R et al (2024) Ferric carboxymaltose and exercise capacity in heart failure with preserved ejection fraction and iron deficiency: the FAIR-HFpEF trial. Eur Heart J 45(37):3789–3800
27. Kittipibul V, Mentz RJ (2024) Intravenous iron therapy in heart failure with preserved ejection fraction: how far have we walked? Eur Heart J 45(37):3801–3803
28. Mebazaa A, Davison B, Chioncel O et al (2022) Safety, tolerability and efficacy of up-titration of guideline-directed medical therapies for acute heart failure (STRONG-HF): a multinational, open-label, randomised, trial. Lancet 400(10367):1938–1952
29. Mebazaa A, Davison B, Chioncel O et al (2023) STRONG-HF Investigators. STRONG-HF: rapid up-titration of guideline-directed medical therapies and intensive follow-up strategy for management of patients hospitalized with acute heart failure. Eur J Heart Fail 25(3):274–282
30. Ter Maaten JM, Davison B, Mebazaa A et al (2024) Safety indicators in patients receiving high-intensity care after hospital admission for acute heart failure: the STRONG-HF trial. J Card Fail 30(2):236–245
31. Arrigo M, Ter Maaten JM, Léopold V et al (2023) Safety, tolerability and efficacy of up-titration of guideline-directed medical therapies for acute heart failure in elderly patients: a sub-analysis of the STRONG-HF randomized clinical trial. Eur J Heart Fail. Epub ahead of print.
32. Solomon SD, McMurray JJV, Vaduganathan M et al (2024) Finerenone in heart failure with mildly reduced or preserved ejection fraction. N Engl J Med 391(16):1475–1485

33. Packer M, Zile MR, Kramer CM, SUMMIT Trial Study Group et al (2025) Tirzepatide for heart failure with preserved ejection fraction and obesity. N Engl J Med 392(5):427–437
34. Kittleson MM, Panjrath GS, Bates K, Breathett K, Dixon DL, Januzzi JL Jr, Mohammed SF (2026) Management of heart failure with preserved ejection fraction: 2026 ACC expert consensus decision pathway: a report of the American College of Cardiology Solution Set Oversight Committee. J Am Coll Cardiol. Epub ahead of print.
35. Kittleson MM, Panjrath GS, Amancherla K et al (2023) 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. J Am Coll Cardiol 81(18):1835–1878
36. US Food and Drug Administration (2026) Biomarker Qualification Program. https://www.fda.gov/drugs/drug-development-tool-ddt-qualification-programs/biomarker-qualification-program. Accessed 30 Jul

Figures

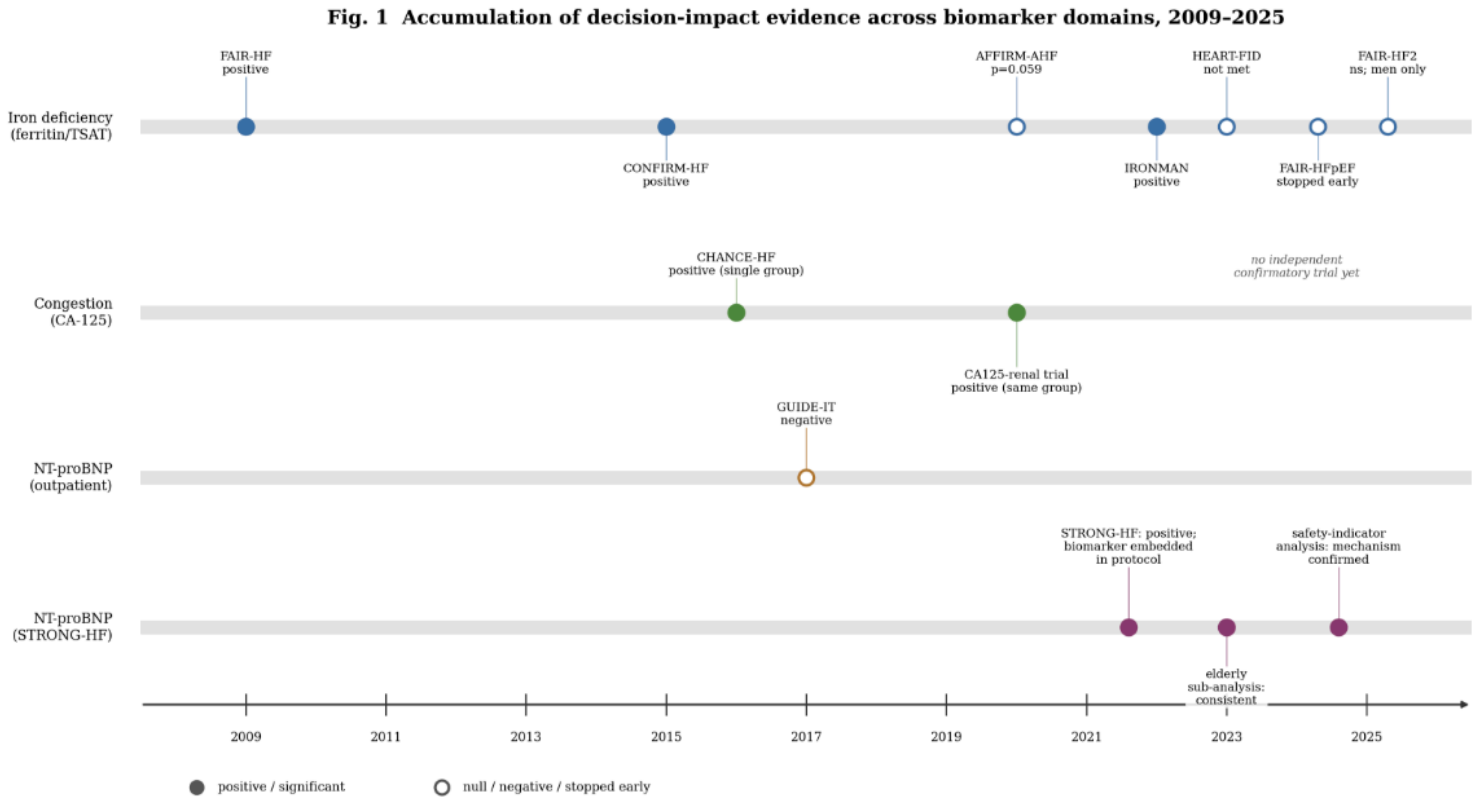


Figure 1

Accumulation of decision-impact evidence across biomarker domains, 2009–2025. Unlike Table 1, which classifies each biomarker's current status, this figure shows how that status was reached over time. Filled markers indicate a positive or significant result; open markers indicate a null, negative, or early-terminated result. The iron-deficiency lane illustrates an inconsistent 16-year pattern rather than steady

accumulation of confirmatory evidence; the CA-125 lane shows two positive but non-independent trials with no subsequent multicenter confirmation; the STRONG-HF lane is the only one showing both an initial positive trial and independent confirmatory sub-analyses using the same embedded biomarker rule.