

METHODOLOGY CORNER

Design and Architecture of a Generative-AI-Supported, Nonphysician-Delivered Model for GDMT Optimization in HFrEF



The ASSIST-HF Trial

Eliano P. Navarese, MD, PhD,^{a,b,c} Joshua H. Leader, MB, ChB, BSc,^{d,e} Raffaella I.L. Markides, BA (HONS),^f Sogol Koolaji, MD, MPH,^{d,g} Dean J. Kereiakes, MD,^h Jacek Kubica, MD, PhD,^c Mehriban Isgender, MD,ⁱ Thomas F. Lüscher, MD,^{e,j,k,*} Diana A. Gorog, MD, PhD^{d,e,g,j,k,*}

ABSTRACT

Patients with heart failure with reduced ejection fraction require rapid initiation and uptitration of guideline-directed medical therapy (GDMT), which is resource-intensive. In a prospective, open-label pilot trial, we assessed the feasibility, acceptability, and safety of a generative artificial intelligence-powered virtual assistant (VA), with retrieval-augmented generation and expert prompt engineering, to optimize GDMT. Patients with new heart failure with reduced ejection fraction (n = 60) were randomized to VA-guided care, delivered by nonmedical staff at 2-weekly intervals or standard-of-care treatment delivered by doctors or nurses. At 12 weeks, patients in the VA arm had superior GDMT optimization across all medication classes, lower N-terminal pro-B-type natriuretic peptide, and fewer hospitalizations. Patient-reported acceptability, appropriateness and feasibility scores were high, with no safety disagreements between VA and clinician recommendations. Treatment by an artificial intelligence-powered VA, *run by nonmedical staff*, with minimal remote medical supervision, is acceptable to patients, and can safely and effectively optimize GDMT, representing a scalable strategy to optimize treatment and health care resource utilization. (JACC Adv. 2026;5:102588) © 2026 The Authors. Published by Elsevier on behalf of the American College of Cardiology Foundation. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

From the ^aDepartment of Life and Health Sciences, Link Campus University, Rome, Italy; ^bSIRIO MEDICINE Research Network, Bydgoszcz, Poland; ^cDepartment of Cardiology and Internal Medicine, Nicolaus Copernicus University, Bydgoszcz, Poland; ^dDepartment of Cardiology, East and North Hertfordshire NHS Trust, Stevenage, Hertfordshire, United Kingdom; ^eFaculty of Medicine, National Heart and Lung Institute, Imperial College, London, United Kingdom; ^fSchool of Clinical Medicine, University of Cambridge, Cambridge, United Kingdom; ^gSchool of Life and Medical Sciences, Postgraduate Medical School, University of Hertfordshire, United Kingdom; ^hThe Christ Hospital and Lindner Research Center, Cincinnati, Ohio, USA; ⁱRepublican Clinical Hospital, Department of Cardiology, Azerbaijan Medical University, Department of Family Medicine, Baku, Azerbaijan; ^jSchool of Cardiovascular and Metabolic Medicine & Sciences, Kings College London, London, United Kingdom; and the ^kRoyal Brompton and Harefield Hospitals, Gy, s and St. Thomas' NHS Foundation Trust, London, United Kingdom. *These authors contributed equally to this work as joint senior authors.

The authors attest they are in compliance with human studies committees and animal welfare regulations of the authors' institutions and Food and Drug Administration guidelines, including patient consent where appropriate. For more information, visit the [Author Center](#).

Manuscript received December 22, 2025; accepted December 23, 2025.

**ABBREVIATIONS
AND ACRONYMS**

AI	= artificial intelligence
BP	= blood pressure
GCP	= Good Clinical Practice
GDMT	= guideline-directed medical therapy
HF	= heart failure
HFrEF	= heart failure with reduced ejection fraction
KCCQ-12	= the Kansas City Cardiomyopathy Questionnaire
LLM	= Large Language Model
NLP	= natural language processing
NT-proBNP	= N-terminal pro-B-type natriuretic peptide
RAG	= retrieval-augmented generation
SOC	= standard of care
VA	= virtual assistant

Key feasibility findings from the ASSIST-HF SIRIO (Safety of AI-Powered Virtual Assistant in Outpatient Management of Heart Failure: A Randomized Controlled Pilot Study; ASSIST-HF SIRIO) randomized pilot trial are summarized in the concurrently published JACC Brief Feature.¹ This Methods Companion provides a comprehensive account of the trial design, the technical architecture of the virtual assistant (VA), the human-in-the-loop governance model, the operational workflows, the safety framework, and the statistical analysis plan that underpinned the ASSIST-HF SIRIO trial.

Heart failure (HF) is a leading cause of morbidity and mortality, with an increasing incidence due to an aging population and improved survival rates for cardiovascular disease.² Acute HF accounts for more than 1 million hospital admissions annually, making it a significant burden on health care systems.³ Contemporary cohorts show that patients recently discharged for HF have a 23% risk of death or readmission within 6 months,⁴ a risk that declines exponentially but remains elevated for up to 2 years.^{5,6} For patients with heart failure with reduced ejection fraction (HFrEF), treatment with optimal guideline-directed medical therapy (GDMT) reduces symptoms, improves prognosis, and can prevent recurrent hospitalization. Trial evidence shows that the beneficial effects of GDMT start early after their initiation.^{4,7}

Although such optimal management is critical, the institution of GDMT relies heavily on frequent outpatient visits to assess the response to medications. Consequently, many patients do not receive optimal GDMT in a timely manner, as shown in real-life registries^{6,8} impacting adversely on prognosis. By leveraging technological advancements, there is potential to optimize medications more rapidly and achieve significant cost savings.

Recent advances in artificial intelligence (AI) and digital health offer transformative solutions.⁹ AI-powered VAs, equipped with capabilities like natural language processing (NLP) and machine learning, may offer a novel avenue to support patients with HFrEF in achieving GDMT.¹⁰ We developed an advanced generative-AI-based HFrEF-specific VA, termed SIRIO-HF, based on the latest guidelines^{11,12} and capable of dynamically tailoring therapy recommendations. This technology combines the retrieval of information from international guidelines with

language generation, allowing the VA to provide responses grounded in up-to-date, domain-specific knowledge.

To our knowledge, ASSIST-HF SIRIO is the first randomized clinical trial in medicine in which active pharmacological treatment recommendations for HFrEF were generated by an advanced generative-AI assistant and delivered by nonphysician staff within a supervised, human-in-the-loop framework.

FULL TRIAL DESIGN AND RATIONALE

The ASSIST-HF SIRIO trial was a single-center, prospective, randomized, open-label pilot study conducted at the Lister Hospital, East and North Hertfordshire NHS Trust in the United Kingdom. It was designed to compare a strategy of frequent VA-guided care visits, delivered by nonmedical staff, with standard-of-care (SOC) treatment delivered by medical or nursing staff in patients recently discharged with a new diagnosis of HFrEF. The study (NCT06400927) received ethics committee approval from the United Kingdom Health Research Authority (Ref 24/WA/0086) and the local Research and Development Board and was conducted in accordance with the Declaration of Helsinki and Good Clinical Practice (GCP).

The rationale for employing a supervised generative-AI model stemmed from the urgent clinical need for rapid GDMT optimization in a high-risk population, contrasted with the operational constraints of resource-intensive, clinician-led follow-up. Treatment of HFrEF with optimal GDMT reduces symptoms, improves prognosis, and can prevent recurrent hospitalization but relies heavily on frequent outpatient visits to assess the response to medications as they are introduced and uptitrated.³ This places a significant burden on health care systems, relying on the availability of health care professionals (cardiologists, general physicians, or cardiac nurses) to conduct visits, resulting in variable care. Consequently, many patients do not receive optimal GDMT in a timely manner, as shown in real-life registries, impacting adversely on prognosis and increasing the likelihood of recurrent hospital admissions.⁶ This places a significant burden on health care systems, relying heavily on the availability of doctors/nurses to conduct visits, resulting in variable care, with many patients not receiving optimal GDMT in a timely manner. The National Health System in England is limited in accommodating frequent outpatient visits due to lack of specialist clinic capacity. In 2024, only 46% of

patients hospitalized with HF received the recommended quadruple therapy promptly.¹³ This impacts adversely on prognosis and increases the risk of recurrent HF admissions.

The study was designed to enhance capacity and standardize care by leveraging a VA to generate guideline-concordant recommendations, while ensuring safety through mandatory human oversight. All VA outputs were reviewed and actioned by a supervising cardiologist, preserving clinician authority as the final decision maker.

The trial was conceived as a pilot study to rigorously assess feasibility, safety, and operational performance before consideration of larger trials. A sample size of 60 patients (30 per arm) was deemed sufficient to provide robust data on these primary objectives and to inform the design of future studies. A single-center implementation was chosen to ensure tight control over protocol fidelity, training of nonmedical staff, and centralized clinical oversight during this initial evaluation phase. The study setting was a secondary care hospital with comprehensive cardiology services, including acute cardiac units and specialist HF clinics.

Written informed consent was obtained by research team members trained in GCP. Patients were randomized using a secure web-based randomization service into 1 of 2 treatment arms with randomization occurring either just before or within 7 days of hospital discharge. Each patient was assigned a unique study ID to ensure anonymity.

STUDY POPULATION AND RECRUITMENT. Participants were identified from inpatient wards, including the acute cardiac and medical units. Eligibility criteria required participants to be aged 18 years or older with newly diagnosed HFrEF and NYHA functional class II-IV symptoms at discharge. This focused the intervention on an incident HFrEF cohort at high short-term risk. Key exclusion criteria were prior history of HF, chronic kidney disease stage 4 to 5 or on dialysis, severe comorbidities that could hinder protocol adherence, alternative diagnoses for symptoms such as chronic pulmonary disease, dementia, or cognitive impairment, and current participation in another interventional clinical trial. These criteria were established to minimize clinical confounders and ensure a population able to adhere to the study protocol. Written informed consent was obtained by research team members trained in GCP for research.

WILLINGNESS OF PATIENTS TO BE RANDOMIZED AND SCREEN FAILURES. A total of 75 patients were screened, of whom 66 met eligibility criteria (Figure 1). Nine patients were excluded from

recruitment: 5 due to prior history of HF, 3 due to chronic kidney disease stage 4, and 1 with severe cognitive impairment. Of eligible patients, 91% agreed to take part. Six patients declined participation; 4 cited travel logistics, 1 was unwilling to undergo further medication changes, and 1 declined due to work commitments and a busy lifestyle.

RANDOMIZATION AND BASELINE ASSESSMENT.

Patients were randomized using a secure web-based randomization service into 1 of 2 treatment arms: a VA-guided treatment group or an SOC group, with randomization occurring either just before or within 7 days of hospital discharge (Figure 2). Each patient was assigned a unique study ID to ensure anonymity. Baseline data, including demographics, medical history, and diagnostic test results such as electrocardiogram and echocardiogram, were recorded at hospital discharge, before the first study visit.

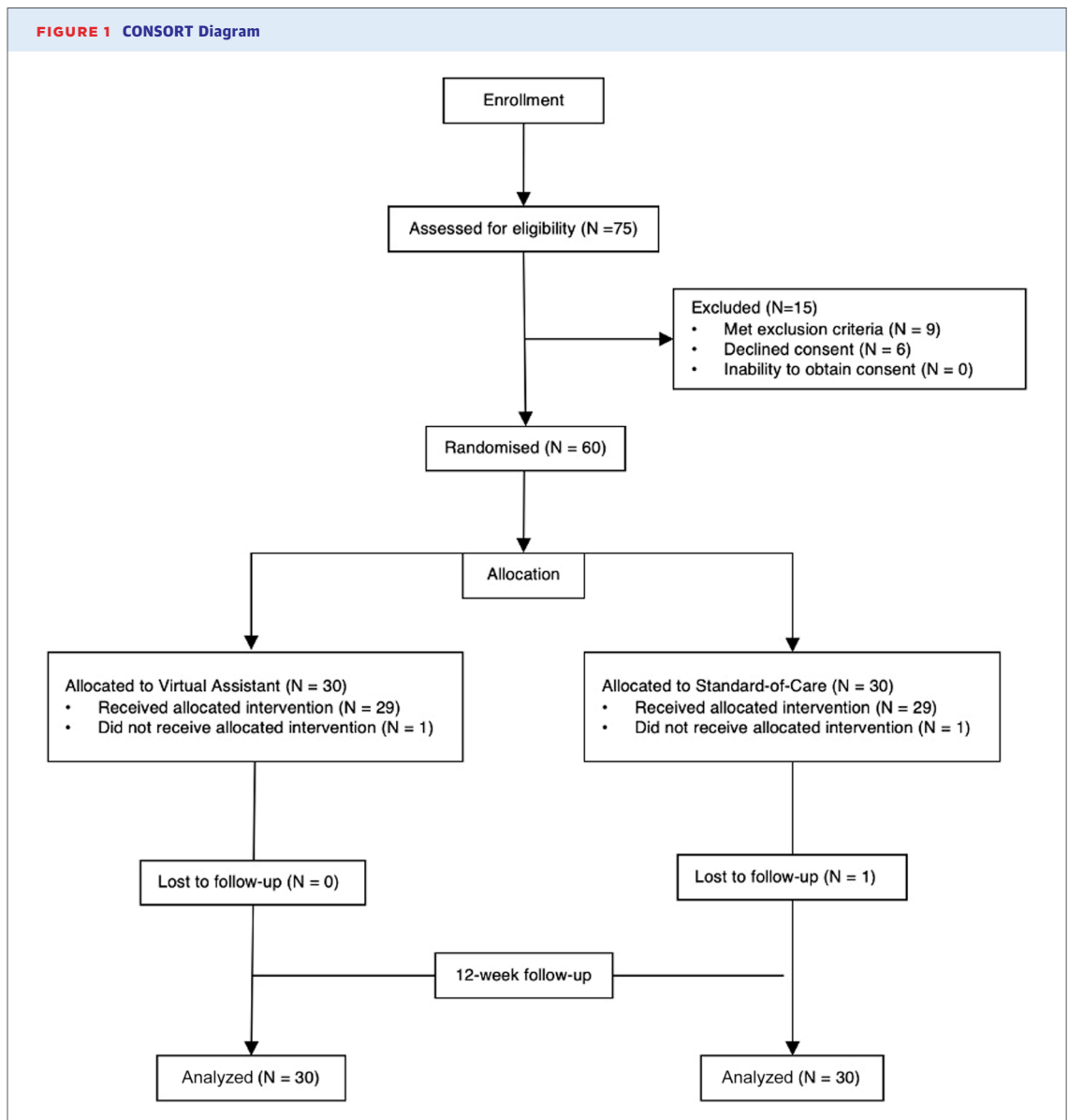
VA ARCHITECTURE

The SIRIO-HF VA was designed as a combined human-AI engine that integrates individual clinical context with structured, evidence-based knowledge. The system employed a large language model (LLM) (GPT-4, OpenAI) coupled with retrieval-augmented generation (RAG), advanced NLP, expert prompt engineering, and systematic clinical validation by cardiologists. The overarching goal was to generate guideline-concordant, personalized GDMT recommendations for patients with HFrEF within a supervised, human-in-the-loop framework (Figure 3).

Recently, ChatGPT (Open AI) was the first generative-AI technology made available to a broad audience that leverages LLM for providing answers to generic questions; this type of tool was trained on vast amounts of data from multiple sources that may lead to incorrect conclusions¹⁴ owing to hallucinations or inaccurate or wrong answers.^{15,16}

The RAG method used in our study searches for and includes relevant, real-world information from selected sources, enhancing AI-generated response with augmented context-based retrieval. This approach has been demonstrated to reduce errors like hallucinations, which are incorrect responses generated by AI when it lacks enough and relevant information.¹⁷⁻¹⁹ LLM enhanced with RAG outperforms standalone LLMs, significantly improving accuracy and relevance, particularly in complex domains such as evidence-based medicine in health care.²⁰

The VA operated on a RAG framework rather than relying solely on the LLM's pretrained internal knowledge. When a visit was initiated, structured and semistructured patient data (symptoms, NYHA

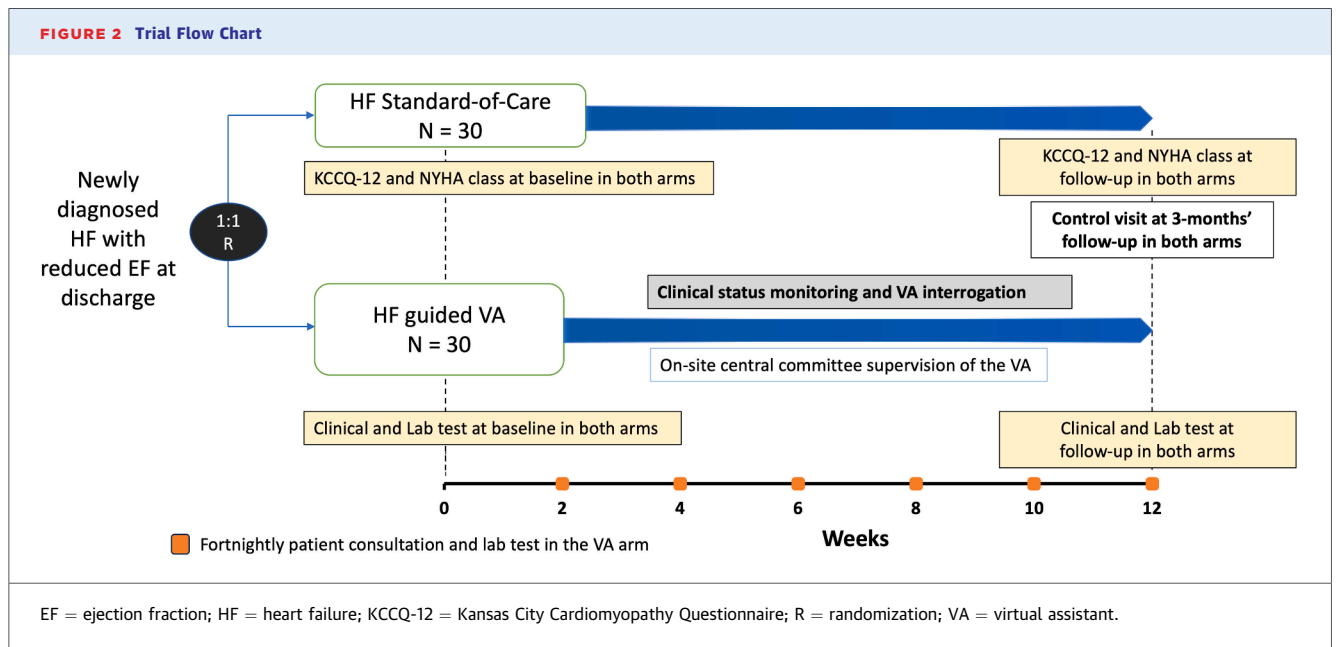


functional class, blood pressure [BP], heart rate, weight, laboratory values, and current HF medications) were encoded and passed through an NLP layer that triggered a targeted retrieval step. This RAG component searched a curated, domain-specific knowledge repository consisting of contemporary HF guidelines and consensus documents.

Retrieved passages were then injected into the LLM prompt, so that all recommendations were explicitly grounded in these up-to-date, cardiology-specific sources rather than in generic web-scale training data. Advanced prompt engineering served as the critical interface between retrieval and response generation: the prompts were designed to:

- 1) force explicit consideration of guideline statements; 2) require the model to justify proposed treatment changes in guideline terms; and 3) restrict outputs to actionable, medication-focused recommendations (initiation, titration, or deferment of GDMT) within predefined safety bounds. This architecture was chosen to mitigate typical LLM limitations such as hallucinations, inconsistent reasoning, and reliance on outdated or noncardiac information, which have been described for general-purpose tools such as ChatGPT trained on heterogeneous data.

In this supervised workflow, nonmedical research staff collected patient data using a standardized script and entered anonymized information into a



secure web interface. The VA generated a structured recommendation for GDMT optimization, which was then reviewed by a remote consultant cardiologist. The physician could fully accept, partially modify, or reject the VA suggestion; in all cases, the clinician’s judgment had absolute priority and the VA functioned strictly as a decision-support tool, not an autonomous prescriber.

DEVELOPMENT PHASE: KNOWLEDGE CURATION AND PROMPT DESIGN. The development phase focused on building a narrow, high-fidelity knowledge base and aligning the LLM’s behavior with HF guideline logic. First, international HF management guidelines and key consensus documents were systematically reviewed. Sections pertaining to HF rEF pharmacotherapy, titration rules, monitoring requirements, contraindications, and special populations were extracted and organized into a structured, machine-readable repository. Appendices, dosing tables, and algorithmic flowcharts were also incorporated to preserve the operational nuance of recommendations.

Cardiologists designed and iteratively refined the prompt templates that govern the VA’s behavior. Prompts specified the following.

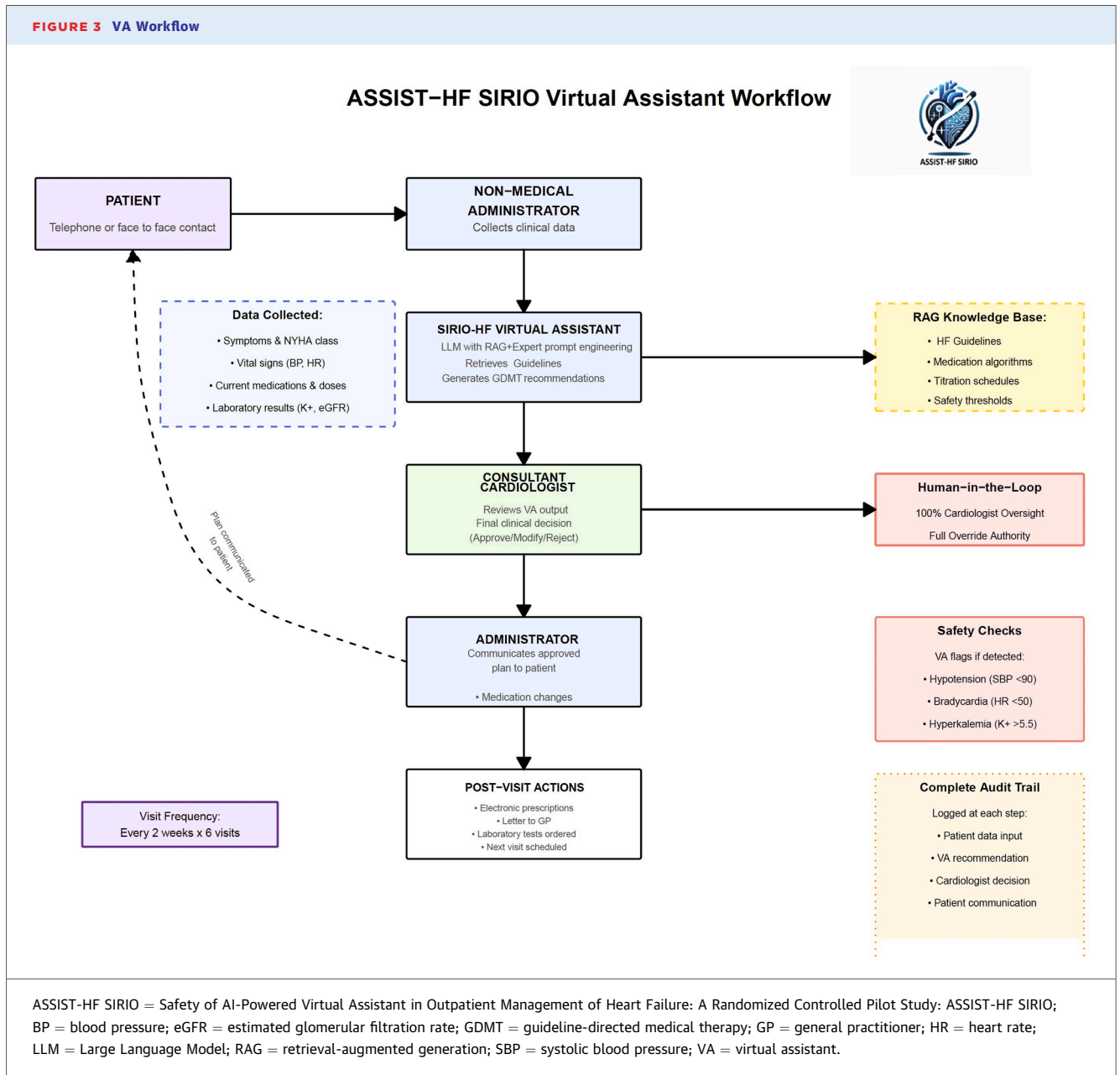
- The required input structure (vital signs, laboratory parameters including glomerular filtration rate and potassium, current GDMT class and dose, recent changes, symptoms, and NYHA functional class)

- The expected output format (explicit class-by-class recommendations, with rationale and priority ordering)
- Hard constraints (for example, not initiating or up-titrating specific agents below predefined heart rate, BP, or estimated glomerular filtration rate (eGFR) thresholds, or in the presence of hyperkalemia)
- An explicit instruction to base all recommendations on the retrieved guideline text and to avoid speculation outside the evidence base.

Through this process, the VA was tuned to behave as a guideline-anchored assistant embedded in an HF clinic workflow rather than as a general conversational agent.

VALIDATION PHASE: REAL-WORLD CASE TESTING AND REFINEMENT. Before being deployed in the randomized trial, the VA underwent a multistage validation process based on hundreds of real-world clinical cases. Anonymized cases representing the spectrum of HF rEF severity, comorbidities, renal function, BP profiles, and medication histories were fed into the system. For each case, the VA generated GDMT recommendations that were independently reviewed by a cardiologist.

Discrepancies between the VA and expert judgment were analyzed in detail. When disagreements reflected incomplete retrieval, ambiguous prompts, or overly permissive logic, the prompts and, where necessary, the retrieval filters and safety thresholds



were revised. This “prompt-correction” feedback loop was repeated iteratively until the steering cardiologists judged the VA’s behavior to be consistently safe and guideline-concordant for routine use within the trial.

SAFETY LOGIC, RED-FLAG DETECTION, AND OVERSIGHT. Patient safety was prioritized by embedding explicit safety logic at multiple levels. Within the VA, red-flag criteria were encoded for symptoms (eg, resting dyspnea, orthopnea, syncope, chest pain), vital signs (eg, systolic BP below a

predefined threshold, bradycardia, oxygen desaturation), and key laboratory parameters (eg, elevated potassium, significant decline in eGFR). When red flags were detected, the VA was constrained to withhold uptitration of GDMT and to recommend clinical escalation rather than pharmacological intensification.

Operationally, every VA output was reviewed in real time by a supervising cardiologist who was not present in the room and did not interact directly with the patient during the visit. Nonmedical staff were trained to escalate immediately if any concerning

symptoms emerged during the interview. All disagreements between the VA and the supervising cardiologist, and any adverse events potentially related to VA suggestions, were logged and periodically reviewed by the trial safety and steering committee, with the option to further tighten prompts or thresholds if needed.

Through this combination of guideline-grounded RAG, expert-designed prompt constraints, pre-deployment case validation, and continuous human oversight, the SIRIO-HF VA was configured as a supervised, safety-first generative-AI assistant intended to support, rather than replace, clinician-led HF management.

Supplemental Figure 1 illustrates a worked example showing data input from research administrator and the VA output for recommended treatment optimization.

USE OF THE VA AND HUMAN-IN-THE-LOOP GOVERNANCE

The VA system, accessed via a secure web interface, processed anonymized clinical data entered by the research administrator. Based on these inputs, the VA generated real-time recommendations for guideline-directed medication optimization (**Supplemental Figure 1**). The SIRIO-HF VA was trained using the latest international guidelines. During consultations, it processed anonymized clinical data, generating therapeutic recommendations. These features distinguished the VA arm from the SOC arm, which relied solely on health care professionals for decision making and care delivery. The VA evaluated these data (real time) and made immediate recommendations to optimize the medications based on the clinical and laboratory status inputted, at the current visit.

The research administrators had no specialist medical/nursing knowledge. They were trained on the principles of good practice in clinical research (GCP). They received simple training on the study, including conventional grading of symptoms by NYHA functional class, questionnaires including patient acceptability and quality-of-life questionnaires, as well as the importance of identification of “red-flag” symptoms (Study Protocol in the **Supplemental Appendix**) and prompt escalation to the supervising clinician. The research assistants were provided with the most recent blood results of the patient, obtained just before each visit. At each visit, the research administrator sense-checked the VA recommendations with the supervising consultant cardiologist (who was not engaged in the consultation nor in the clinic room) to ensure the clinician agreed with the

VA recommendations. This hierarchical decisional approach ensured that the therapeutic recommendations made by the VA were safe and appropriate for each patient.

If the clinician approved the VA recommendations, the research administrator advised the patient of these recommendations. For any VA-recommended changes in medications, prescriptions were issued by the clinician in charge (cardiologist), who had no contact with the patient. Following the consultation, the agreed management plan was communicated in an outcome letter with the updated medical status and recommendations, issued to the patient by the researcher, with a copy sent to the primary care physician. If the supervising clinician did not agree wholly or in part with the VA recommendations, the clinician made the ultimate decision and the clinician’s decision always over-rode the VA.

OPERATIONAL WORKFLOW AND PROCEDURES

The operational workflow for each VA-supported visit was highly detailed and standardized. Patients in this arm attended visits at 2-weekly intervals, either in person or via telephone. During the visit, the administrator used a script to gather information on symptoms, vital signs, and medications, which was then entered into the VA’s secure web interface. The VA processed this information and generated a recommendation, which was immediately routed to the supervising cardiologist for review. Once the clinician finalized the plan, it was communicated to the patient, prescriptions were issued, and a letter was sent to the patient and their primary care physician. In contrast, patients in the SOC arm followed the usual HF management pathway at the institution. This involved outpatient consultations with doctors or HF nurse specialists, with the frequency and modality of visits determined by clinical need and service availability, reflecting real-world practice.

VA ARM. Patients in the VA arm were treated with the integration of the SIRIO-HF VA into their management. Study visits were conducted either in person or via the telephone, according to the patient’s wishes. Visits were conducted at 2-weekly intervals by nonmedical researchers (who had no medical or nursing qualifications) and who were simply instructed to ask scripted questions and collate responses. Patients were asked to attend for a blood test through the conventional hospital phlebotomy service, ideally 1 to 2 days before the research visits. The results of the blood tests were made available to

the research administrator for each visit. Lab tests including urea and electrolytes, creatinine, eGFR, N-terminal pro-B-type natriuretic peptide (NT-proBNP), and full blood count were assessed every 2 weeks as needed (just before the next visit), in line with GCP during HF medication optimization.

At each visit, a trained research administrator assessed the health status of the patient (update), compared with that at the last contact (or compared with the status at hospital discharge, in the case of the first research visit), which began with a review of the patient's symptoms and overall well-being according to a standardized set of questions (Study Protocol in the [Supplemental Appendix](#)), including NYHA functional class, BP, heart rate, and weight.

At the end of each visit, the research administrator imported this information, together with the latest blood results, in real time, onto the VA web interface, in anonymized format. This yielded an output of VA recommendations for personalized optimized GDMT and the researcher passed on this recommendation to the supervising cardiologist. Access to the interface was login and password protected, and available only to the research team. At the beginning and at the end of the study, the Kansas City Cardiomyopathy Questionnaire (KCCQ-12) and a Patient Satisfaction Survey were administered by the research team.

At the end of the study, patients in the VA arm were transferred to continue SOC follow-up, outside of the study, conducted either by a doctor or an HF nurse specialist.

SOC ARM. Patients in the SOC arm followed the usual HF management pathway, which involved regular outpatient consultations conducted by doctors or HF nurse specialists, either face-to-face or remotely via the telephone, according to the patient's preference and availability of appointments.

Patients are normally recommended to purchase a home BP monitor and to discuss home BP and heart rate readings at the visits, or alternatively the BP is measured by the health care professional at each visit, if conducted face-to-face. Patients are also advised to weigh themselves regularly at home and to discuss their weight at each clinic visit. Patients attending face-to-face consultations are weighed by the health care professional. During outpatient optimization of medication, blood tests are frequently performed including urea and electrolytes, creatinine, glomerular filtration rate, NT-pro-brain natriuretic peptide (NT-proBNP) and full blood count. At each visit, clinical symptoms and medication changes were logged, as normal practice. Quality of

life was assessed at the beginning and end of the study using the KCCQ-12 and a Patient Satisfaction Survey.

PATIENT ACCEPTABILITY AND FEASIBILITY. At the end of the study, a modified Acceptability of Intervention Measure, Intervention Appropriateness Measure, and Feasibility of Intervention Measure scores with visual analogue scales were administered by the research team to patients in the VA arm ([Supplemental Figure 2](#)).

PROTOCOL AND STATISTICAL ANALYSIS PLAN

The full trial protocol can be found in the [Supplemental Appendix](#). The primary outcomes of the trial were feasibility, acceptability, and the effectiveness of the VA-guided system in optimizing GDMT over 12 weeks compared with SOC. Secondary outcomes included changes in NYHA functional class, NT-proBNP levels, rates of HF-related hospitalizations, and patient-reported quality of life as assessed by the KCCQ-12. All primary efficacy analyses were conducted on an intention-to-treat basis. Given the pilot nature of the trial, *P* values were considered exploratory.

A sample size of 60 patients (30 per group) was estimated to provide adequate data for feasibility and safety assessments and exploratory analysis of secondary outcomes. This sample size was based on practical feasibility considerations rather than formal power calculations due to the primary focus on feasibility.

The study design and analysis followed the intention-to-treat principle. Any individuals withdrawn, either through patient or clinician decision, were transferred to SOC, but data already collected were used and follow-up data collected if agreed by the patient.

Data distribution was checked by using the Kolmogorov-Smirnov test with data presented as median with IQR or mean \pm SD as appropriate. Categorical variables are summarized as percentages. Comparisons between treatment groups (VA vs SOC) were conducted to evaluate the optimization of GDMT across 4 key medication classes: angiotensin-converting enzyme inhibitors/angiotensin receptor blockers (ACEI/ARB), angiotensin receptor neprilysin inhibitors (ARNI), beta-blockers (BB), and mineralocorticoid receptor antagonists (MRA), as well as uptake of sodium-glucose co-transporter 2 inhibitor (SGLT2i). The proportions of patients categorized by GDMT dose level (full optimal dose, half to less than

full optimal dose, less than half optimal dose, or no treatment) were summarized as percentages at baseline and end-of-treatment. For each medication class (ACEI/ARB/ARNI, BB, MRA, and SGLT2i), the proportion of patients reaching maximal tolerated doses was calculated by treatment group. Patient acceptability of the VA was assessed through 3 validated questionnaires, with the average score calculated across the scales. Acceptability was summarized descriptively, reporting the proportion of patients scoring above a predefined threshold of 80%.

A timeline analysis was conducted to illustrate the proportion of patients achieving full optimal dose at each study visit, summarized descriptively over time by treatment group (VA vs SOC).

Secondary endpoints, including functional class and quality of life, were recorded as absolute values and change compared with baseline within group and between the VA and the SOC groups. NT-proBNP levels were log-transformed. Continuous variables were inspected for distributional assumptions; for skewed variables such as NT-proBNP, values were analyzed on the log scale. Between-group comparisons were performed using Student's *t*-test on log-transformed values after verifying homogeneity of variances.

PARTICIPANT FLOW AND BASELINE CHARACTERISTICS

The flow of participants through the trial is detailed in the CONSORT (Consolidated Standards of Reporting Trials) diagram ([Figure 1](#)) and the study flow chart is shown in [Figure 2](#). Sixty patients with HFREF were enrolled and randomized 1:1 to either VA-guided or SOC outpatient management over a 12-week period. The VA-guided arm was delivered by nonmedical/non-nursing staff, overseen remotely by a doctor and the SOC arm was delivered by doctors and nurses in line with usual care.

This AI randomized controlled trial follows the consensus statement for clinical trial reports for interventions involving AI, the CONSORT, and CONSORT-AI extension.^{21,22} The 14 new items are reported to promote transparent reporting of AI intervention and are intended to facilitate critical appraisal and evidence synthesis ([Supplemental Table 1](#)).

Baseline characteristics, including demographics, medical history, and diagnostic test results, were recorded at hospital discharge for all randomized patients and were generally balanced between the study arms.

SAFETY FRAMEWORK

Patient safety was paramount. The study protocol included a highly detailed safety framework with predefined thresholds for vital signs and laboratory values that would trigger specific actions, such as withholding medication or escalating care. An emergency pathway was established for any patient experiencing a substantial worsening of symptoms; in such cases, the research protocol would be bypassed, and the supervising clinician would make appropriate clinical decisions, including hospital admission if necessary. All disagreements between the VA and the clinician were logged and reviewed by the steering committee to identify opportunities for model refinement and to ensure ongoing safety.

SAFETY PROCEDURES AND OVERSIGHT. A trial safety and steering committee was established to oversee the conduct and the safety of the study. To establish a standardized process for addressing and resolving disagreements between the supervising physician's clinical judgment and the therapeutic recommendations made by the VA, a protocol of agreement was established. Data regarding agreement between the VA and the clinician at each visit, for each patient, were collected and closely monitored. Although decision making was real time, all cases of disagreement between the VA and the consultant were reviewed by the steering committee. Adverse events were categorized as either: 1) therapeutic inappropriateness of VA suggestions; or 2) technical issues with the VA that could affect patient care, like system errors or misinformation.

If a patient in the VA arm were to experience substantial worsening of HF symptoms or reported symptoms of concern that might require hospitalization, based on "red-flag" warning signs (Study Protocol in the [Supplemental Appendix](#)) detected by the research administrator or the VA, the clinician supervising would be informed and would make appropriate clinical decisions (including admission to hospital) as needed, outside of the research protocol.

The study design followed the intention-to-treat principle. Any individuals withdrawn, either through patient or clinician decision, were transferred to SOC, but data already collected were used and follow-up data collected if agreed by the patient.

MECHANISTIC OPERATIONAL EXPLANATION

The architecture of the VA-supported workflow was designed to support GDMT optimization through several key operational mechanisms. First, the enforced 2-weekly visit cadence ensures regular and

frequent opportunities for medication review, reducing the risk of therapeutic inertia. From a patient perspective, knowing that they have a visit scheduled in the near future gives patients confidence to follow the recommended treatment uptitration, knowing they will soon have a touchpoint to discuss any concerns or side effects. Second, by grounding its recommendations in a curated knowledge base of international guidelines, the VA promotes the consistent application of evidence-based therapy, thereby reducing interclinician variability in titration practices. Third, the systematic collection and evaluation of safety parameters at each visit provides a structured safety net, aiming to prevent inappropriate medication changes. Finally, by delegating data collection to nonmedical staff, the model is designed to increase the capacity of a specialist clinical service, allowing physicians to oversee a larger volume of titration visits and focus their time on more complex clinical decisions.

IMPLEMENTATION AND SCALABILITY ROADMAP

The findings from this pilot study inform a potential roadmap for broader implementation. Scaling this supervised AI-supported model to larger systems would require a robust and secure digital infrastructure. Standardized data input is essential and so training administrative staff to ask and record data in a standardized comprehensive format would be essential. The ratio of administrators to supervising clinicians would need to be optimized to balance efficiency with safe oversight. Although the model's core components are adaptable, implementation in different health care systems would require tailoring to local guidelines, staffing models, and regulatory environments. A robust governance framework, with clear lines of accountability and continuous performance monitoring, would be critical for any large-scale rollout. By potentially alleviating the burden on specialist HF clinics, this model could allow clinicians to dedicate more time to patients with the most complex needs. In light of the significance of the disease, noting the global burden and significant impact of HF, and potential benefit of enhanced and complete treatment escalation including in underserved populations, application of this technology may be particularly important in under-resourced parts of the world where there is reduced access to specialist care. Health economic evaluation would be essential in different settings, to ensure the model is suitably funded and aligned with different health

care systems. Finally, scalability would need to consider the acceptability of this technology to patients who may have different socioeconomic, cultural, educational, and health literacy levels.

DISCUSSION

We show that frequent VA-guided management, administered by nonmedical staff, of high cardiovascular risk outpatients to optimize medication, is feasible, acceptable to patients, and achieves at least as good, if not better outcomes at 12 weeks compared with SOC visits with medical staff.

As HF remains a leading cause of global mortality and hospitalizations, timely institution of optimal GDMT is pivotal to improving clinical outcomes, reducing hospitalizations, and improving patient quality of life.²³⁻²⁵ Currently, GDMT adherence in the real world remains problematic, particularly regarding therapeutic optimization and medication titration.^{26,27} The integration of AI into GDMT optimization represents a novel avenue, yet existing data are limited, mostly derived from traditional predictive models and simple decision checklists.²⁸⁻³¹

This novel AI-guided HF management randomized trial addresses the clinical value of a generative-AI tool to enhance GDMT, administered by nonmedics. This randomized trial demonstrated that using the AI-driven VA delivered by nonmedical personnel under physician oversight enhanced GDMT adherence and therapy optimization.

Patient safety was rigorously prioritized through specialized model training for symptom red-flag detection, continuous safety, and oversight by expert cardiologists, ensuring adherence to robust clinical safety standards.

STUDY LIMITATIONS. This was a pilot trial; therefore, no formal sample size calculations have been performed. This was a single-center study conducted in England. The SOC patient management in the National Health Service is limited by availability of specialist clinics and cardiac services in proportion to needs. Therefore, the SOC comparison may not be directly translatable to other health care systems. In addition, the demographics, ethnicity, educational level and literacy, and approach to AI of the patient cohort may be different from that of other populations. All findings should be interpreted as hypothesis generating. Further large powered clinical trials are needed to investigate the potential of AI-powered VA in HF management. The VA responses were reviewed by an expert physician cardiologist;

thus, this tool should be interpreted as an assistant aiming to support but not replace human clinical judgment.

CONCLUSIONS

Integration of an AI-driven VA, delivered by nonmedical staff, into the care of patients with newly diagnosed HFrEF, with minimal remote clinician oversight, improved GDMT uptake and optimization, reduced HF-related hospitalizations and biomarkers, and achieved high patient acceptance. This hybrid strategy may represent an effective, patient-centered approach to enhance HF outpatient management and optimize health care resources.

ACKNOWLEDGMENTS The authors are very grateful to the nonmedical and administrative staff for their support in facilitating the study.

FUNDING SUPPORT AND AUTHOR DISCLOSURES

Dr Gorog has received institutional research grants from Medtronic and AstraZeneca, Speakers Bureau for Chiesi; and advisory board fees from Janssen and BMS, all outside of the submitted work. Dr Lüscher has received institutional educational grants from Abbott, Amgen, AstraZeneca, Boehringer Ingelheim, Daichi-Sankyo, Novartis, Eli Lilly, Novo Nordisk, Sanofi, Vifor, and Bayer; and consulting fees from Milestone Pharmaceuticals and Novo Nordisk, all outside of the submitted work. All other authors have reported that they have no relationships relevant to the contents of this paper to disclose.

ADDRESS FOR CORRESPONDENCE: Dr Diana A. Gorog, Faculty of Medicine, National Heart and Lung Institute, Imperial College, Dovehouse Street, London SW3 6LY, United Kingdom. E-mail: d.gorog@imperial.ac.uk. OR Dr Eliano P. Navarese, Department of Life and Health Sciences, Link Campus University, Via del Casale di S. Pio V, 44, Rome 00165, Italy. E-mail: elianonavarese@gmail.com. X handle: [@ElianoNavarese](https://x.com/ElianoNavarese).

REFERENCES

- Navarese EP, Leader JH, Markides RIL, et al. AI-guided GDMT optimization after HFrEF hospitalization: the ASSIST-HF SIRIO randomized pilot trial. *J Am Coll Cardiol*. 2026. <https://doi.org/10.1016/j.jacc.2025.12.066>
- Groenewegen A, Rutten FH, Mosterd A, Hoes AW. Epidemiology of heart failure. *Eur J Heart Fail*. 2020;22:1342-1356.
- Agarwal MA, Fonarow GC, Ziaeian B. National trends in heart failure hospitalizations and readmissions from 2010 to 2017. *JAMA Cardiol*. 2021;6:952-956.
- Mebazaa A, Davison B, Chioncel O, et al. 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*. 2022;400:1938-1952.
- Solomon SD, Dobson J, Pocock S, et al. Influence of nonfatal hospitalization for heart failure on subsequent mortality in patients with chronic heart failure. *Circulation*. 2007;116:1482-1487.
- Metra M, Adamo M, Tomasoni D, et al. Pre-discharge and early post-discharge management of patients hospitalized for acute heart failure: a scientific statement by the Heart Failure Association of the ESC. *Eur J Heart Fail*. 2023;25:1115-1131.
- Zheng J, Sandhu AT, Bhatt AS, et al. Inpatient use of guideline-directed medical therapy during heart failure hospitalizations among community-based health systems. *JACC: Heart Fail*. 2025;13:43-54.
- Dunlay SM, Eveleth JM, Shah ND, McNallan SM, Roger VL. Medication adherence among community-dwelling patients with heart failure. *Mayo Clin Proc*. 2011;86:273-281.
- Bajwa J, Munir U, Nori A, Williams B. Artificial intelligence in healthcare: transforming the practice of medicine. *Futur Healthc J*. 2021;8:e188-e194.
- Ambrosy AP, Fonarow GC, Butler J, et al. The global health and economic burden of hospitalizations for heart failure. *J Am Coll Cardiol*. 2014;63:1123-1133.
- McDonagh TA, Metra M, Adamo M, et al. 2023 Focused update of the 2021 ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure. *Eur J Heart Fail*. 2024;26:5-17.
- McDonagh TA, Metra M, Adamo M, et al. 2021 ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure. *Eur J Heart Fail*. 2022;24:4-131.
- National Institute for Cardiovascular Outcomes Research. National Cardiac Audit Programme. National Heart Failure Audit (NHFA): 2025 summary report. 2025. London: Healthcare Quality Improvement Partnership (HQIP). 2025. <https://www.nicor.org.uk/interactive-reports/national-heart-failure-audit-nhfa>. Accessed September 01, 2025.
- Zhang J, Zhang Z. Ethics and governance of trustworthy medical artificial intelligence. *BMC Méd Inform Decis Mak*. 2023;23:7.
- Rydzewski NR, Dinakaran D, Zhao SG, et al. Comparative evaluation of LLMs in clinical oncology. *NEJM AI*. 2024;1(5). <https://doi.org/10.1056/aioa2300151>
- Singhal K, Azizi S, Tu T, et al. Large language models encode clinical knowledge. *Nature*. 2023;620:172-180.
- Biesheuvel LA, Workum JD, Reuland M, et al. Large language models in critical care. *J Intensiv Med*. 2025;5:113-118.
- Nolin-Lapalme A, Theriault-Lauzier P, Corbin D, et al. Maximising large language model utility in cardiovascular care: a practical guide. *Can J Cardiol*. 2024;40:1774-1787.
- Zakka C, Shad R, Chaurasia A, et al. Almanac—Retrieval-augmented language models for clinical medicine. *NEJM AI*. 2024;1(2). <https://doi.org/10.1056/aioa2300068>
- Woo JJ, Yang AJ, Olsen RJ, et al. Custom large language models improve accuracy: comparing retrieval augmented generation and artificial intelligence agents to noncustom models for evidence-based medicine. *Arthroscopy*. 2025;41:565-573.e6.
- Schulz KF, Altman DG, Moher D, Fergusson D. CONSORT 2010 changes and testing blindness in RCTs. *Lancet*. 2010;375:1144-1146.
- Liu X, Rivera SC, Moher D, et al. Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: the CONSORT-AI extension. *Nat Med*. 2020;26:1364-1374.
- Fonarow GC, Yancy CW, Hernandez AF, Peterson ED, Spertus JA, Heidenreich PA. Potential impact of optimal implementation of evidence-based heart failure therapies on mortality. *Am Heart J*. 2011;161:1024-1030.e3.
- Fonarow GC, Hernandez AF, Solomon SD, Yancy CW. Potential mortality reduction with optimal implementation of angiotensin receptor neprilysin inhibitor therapy in heart failure. *JAMA Cardiol*. 2016;1:714-717.
- Tang AB, Ziaeian B, Butler J, Yancy CW, Fonarow GC. Global impact of optimal implementation of guideline-directed medical therapy in heart failure. *JAMA Cardiol*. 2024;9:1154-1158.
- Greene SJ, Butler J, Albert NM, et al. Medical therapy for heart failure with reduced ejection

fraction the CHAMP-HF Registry. *J Am Coll Cardiol*. 2018;72:351-366.

27. Fonarow GC, Albert NM, Curtis AB, et al. Improving evidence-based care for heart failure in outpatient cardiology practices: primary results of the Registry to Improve the Use of Evidence-Based Heart Failure Therapies in the Outpatient Setting (IMPROVE HF). *Circulation*. 2010;122:585-596.

28. Yasmin F, Shah SMI, Naeem A, et al. Artificial intelligence in the diagnosis and detection of heart failure: the past, present, and future. *Rev Cardiovasc Med*. 2021;22:1095-1113.

29. Schuurin MJ, Treskes RW, Castiello T, et al. Digital solutions to optimize guideline-directed medical therapy prescription rates in

patients with heart failure: a clinical consensus statement from the ESC Working Group on e-Cardiology, the Heart Failure Association of the European Society of Cardiology, the Association of Cardiovascular Nursing & Allied Professions of the European Society of Cardiology, the ESC Digital Health Committee, the ESC Council of Cardio-Oncology, and the ESC Patient Forum. *Eur Heart J Digit Heal*. 2024;5:670-682.

30. Allen LA, Venechuk G, McIlvennan CK, et al. An electronically delivered patient-activation tool for intensification of medications for chronic heart failure with reduced ejection fraction. *Circulation*. 2021;143:427-437.

31. Kim R, Suresh K, Rosenberg MA, et al. A machine learning evaluation of patient charac-

teristics associated with prescribing of guideline-directed medical therapy for heart failure. *Front Cardiovasc Med*. 2023;10:1169574.

KEY WORDS artificial intelligence, GDMT, heart failure

APPENDIX To ensure full operational transparency, the Supplemental Appendix includes the complete study protocol, a detailed description of the ethical and privacy safeguards implemented in the trial, supplemental figures, and a supplemental table. To view the supplemental appendix, please see the online version of this paper.