Title: The problem with hospitalization endpoints in heart failure trials.

Article type: Viewpoint.

Authors: Sam Straw[1]* MBChB, PhD, Pieter Martens[2] MD, MSc, PhD, Nathan Mewton[3] MD, PhD, Klaus K Witte[1] MBBS, MD, Wilfried Mullens[3] MD, PhD.

Affiliations: [1]University of Leeds, Leeds, UK. [2]Ziekenhuis Oost-Limburg, Genk, and University Hasselt,Belgium. [3]Université Claude Bernard Lyon 1, Lyon, France.

Corresponding author: Professor Wilfried Mullens. Professor of Cardiovascular Physiology, Hasselt University, Belgium, and Heart Failure and Cardiac Rehabilitation Specialist, Ziekenhuis Oost-Limburg, Genk, Belgium. Phone: (+32) 89327087 Fax: (+32) 89327918 Email: wilfried.mullens@zol.be

Contributions: All authors have contributed equally to the manuscript, and all have reviewed the contents prior to submission.

Funding: None.

Acknowledgements: SS is supported by a NIHR Clinical Lectureship.

Conflicts of interest: SS reports speaker's fees and honoraria from Astra Zeneca. NM reports speaker's fees and honoraria from Astra Zeneca, Novartis, BMS, Amgen, Cardiac Dimensions, Boehringer Ingelheim, Bayer and Novo Nordisk. KKW reports speaker's fees and honoraria from Medtronic, Abbott, Cardiac Dimensions, Novartis, BMS, Pfizer, Bayer and unconditional research grants from Medtronic. None of the other authors have anything to declare.

Ethics committee approval: Not applicable.

Word count: 1734

Key words: Clinical trials, hospitalization, heart failure, methodology.

The changing landscape of heart failure trials

Most early trials for heart failure (HF) assessed the effect of therapies on haemodynamic parameters, functional capacity and symptoms. Methodological limitations of these trials, as well the discovery that improvements in symptoms do not always correlate with better outcomes[1] resulted in the inclusion of cardiovascular safety as a mandated trial endpoint. The near parallel discovery that neurohormonal blockade could improve survival led to a paradigm in which new treatments received regulatory approval on this basis. As prognosis improved with successive therapies, event rates in clinical trials declined, meaning ever-larger sample sizes were required (Graphical Abstract). Further confounding the situation, participant profiles have shifted from a population of younger people with a single condition, to one which is older and increasingly multimorbid. With this has come an increasing contribution of non-cardiovascular events, which dilute the treatment effect of HF specific therapies.[2]

Why hospitalizations matter

A hospitalization for HF is a significant event, commonly followed by a permanent decline in quality-of-life, functional status, and independence. This might result from direct harms of the hospitalization itself or simply reflect the progression of the underlying HF syndrome. Hospitalizations are not only the costliest component of HF care but, given its high and increasing prevalence, also form a large proportion of *overall* healthcare expenditure. Regulatory bodies and guideline committees now consider reducing hospitalizations a valid reason to recommend treatments, and so capturing and reporting these data in HF trials is essential.

So, what's the problem?

Problem 1: Hospitalizations are different from other endpoints

Whilst hospitalizations (and their causes) are more difficult to assess than total mortality, preventing both death and hospitalization are important to patients. Including hospitalization events within composite endpoints is, therefore, a legitimate strategy to provide a broader assessment of treatment effects, whilst also having the advantage of increasing statistical power. Nevertheless, in time-to-first or first-and-recurrent event analyses, hospitalization events are regarded as being equally important to death although intuitively, this makes no sense. Whilst balancing the relative importance of mortality and hospitalization endpoints is challenging, this conflict is limited if the direction of the effect is concordant. Discordant endpoints complicate trial interpretation. In FINEARTS-HF (Finerenone Trial to Investigate Efficacy and Safety Superior to Placebo in Patients with Heart Failure) reductions in worsening HF events were not corroborated by reductions in cardiovascular death or total mortality.[3] Whilst this trial is widely viewed as 'positive' and potentially practice changing, comparable reductions in HF hospitalizations in DIG (Digitalis Investigation Group) have been largely forgotten, perhaps in part due to the choice of all-cause mortality as this trial's primary endpoint.[4]

Hierarchical clinical composite endpoints attempt to take account of how mortality, hospitalization and patient-reported outcomes differ in their relative importance, whilst also increasing statistical power and including more information relevant to patients.[5] However, subjective endpoints may dominate where event rates are low. In trials assessing transcatheter tricuspid edge-to-edge repair for right HF adverse clinical events were uncommon, and so the positive results were driven by unblinded quality-

of-life assessments, which are challenging to assess since most scoring systems are influenced by co-morbidities in this population.[6,7] The conflation of surrogate markers with mortality and hospitalization endpoints within hierarchical designs is another concern. For example, lower than anticipated event rates in DAPA-MI (Dapagliflozin in Patients with Myocardial Infarction) resulted in the amendment of the primary endpoint to a hierarchical clinical composite including 'cardiometabolic' endpoints including the new onset of diabetes or modest reductions in body weight.[8]

Problem 2: Hospitalizations are not always due to heart failure

Many contemporary 'positive' trials report reductions in HF hospitalizations whilst the *overall* burden of hospitalization is unchanged. This has become increasingly common for trials conducted in populations with lower event rates, or competing risks of non-cardiovascular morbidity and mortality, such as those for HF with preserved ejection fraction. In EMPEROR-Preserved (Empagliflozin Trial in Patients with Chronic Heart Failure with Preserved Ejection Fraction), empagliflozin reduced HF hospitalizations, however 33 participants needed to be treated for two years to prevent one admission, and only 18% of total hospitalisations were due to HF.[9] The net benefit to patients and healthcare systems of shifting the cause without reducing overall admissions remains unclear, and patients likely care little whether they are hospitalised for HF or for another reason.

Problem 3: Hospitalizations are complex, inconsistently defined and not exclusively due to heart failure

Hospitalizations are often complex, with overlapping causes and a spectrum of presentations ranging from acute pulmonary oedema to more gradual volume

overload. Yet, trials require binary categorisation: was the admission for HF, or not? The approach most widely adopted is for investigator-reported hospitalization episodes to be adjudicated by committees who impose strict criteria, usually requiring symptoms plus diuretic intensification, which may exclude cases where HF contributed indirectly. These criteria are not standardized and may differ between trials, complicating trial interpretation. Additionally, in the process of adjudication much information is lost and hospitalizations which include worsening HF are precipitated by another cause (such as infection) may be simply ignored. The notion that we should differentiate between spontaneous and precipitated decompensation episodes may be counter-productive, since both are independently associated with increased mortality risk.[10]

Where the effect size of an intervention is large, the exact approach to endpoint accumulation, whether investigator-reported or centrally adjudicated may matter little.[11] However for many trials ignoring potentially relevant events may be detrimental to interpretation. One possible solution might be to approach adjudication in a non-binary fashion and assign a probability as to whether episode is related to the outcome of interest – in other words, instead of treating each event as counting or not, it incorporates the uncertainty of adjudication directly into the analysis. This approach increases statistical power and may affect the interpretation of trials with statistically fragile results.[12] Another solution is to use routinely collected electronic healthcare data as an unbiased approach to endpoint collection. Whilst such routine data might risk misclassifying some events, this does not necessarily result in bias of the estimated treatment effect, so long as inaccuracies are randomly distributed across treatment arms – which should be the case in trials which are double-blinded.

Adjudication is more problematic in open-label trials, where participants and investigators are aware of treatment allocation. If endpoint collection is biased, even a small increase in the number of false events in the control arm can significantly overestimate the true treatment effect.

The logical solution to these limitations is to have adjudication committees blinded to treatment allocation. The assumption is that the committee's lack of direct knowledge of the participant, and the imposition of strict adjudication criteria improves the validity of these decisions. However, committees adjudicate using source data, including medical documentation and discharge letters which have been compiled by unblinded physicians. Therefore, in open-label trials even blinded adjudication does not mitigate against the risk of ascertainment bias, making less bias prone endpoints preferable.

Problem 4: Hospitalization is a decision

Unlike death, the decision to hospitalize a patient is a subjective judgement made by the treating physician. Although for some patients the need for hospitalization might be unequivocal, there may be differences in the reasons for and thresholds to hospitalize a patient between different physicians and healthcare systems. The knowledge of a participant's' treatment allocation has the potential to influence the decision to admit a patient to hospital or manage them in the community, whether consciously or not. This is a particular concern in unblinded trials of medical devices, where physicians may have lower thresholds to admit untreated patients. In RESHAPE-HF-2 (A Randomized Study of the MitraClip Device in Heart Failure Patients With Clinically Significant Functional Mitral Regurgitation) reductions in

hospitalizations and improvements in quality of life were not matched by changes in mortality or objective markers.[13]

One approach to mitigate against the subjective nature of hospitalization decisions might be to include other clinically meaningful events within an endpoint designed to capture worsening HF, such as urgent outpatient visits, or the intensification of diuretic therapy. This approach has been increasingly utilized in contemporary trials and provides additional nuance beyond binary hospitalization episodes.[3] However, the potential downsides of this approach are that these events are even more prone to variability in practice patterns, are less prognostically relevant, and may increase the adjudication burden. All of these are factors could dilute the clinical interpretability of trial results compared to an approach which uses more robust outcomes.

Problem 5: Patients guide hospitalization decisions

Bias is not limited to physicians, as patients aware of their assignment may alter their reporting and behaviour. In medical device trials, those randomised to the control arm are effectively 'denied' an intervention which many would consider as standard care. In EARLY TAVR (Evaluation of TAVR Compared to Surveillance for Patients with Asymptomatic Severe Aortic Stenosis) patients with objectively asymptomatic aortic stenosis randomized to usual care showed early and unexpected increases in hospitalizations,[14] likely due to increased vigilance about their untreated disease, so called 'subtraction anxiety'.[15] Drug trials are less prone to this form of bias, but where the effects are obvious, blinding can still be compromised. Tirzepatide, for instance, causes rapid weight loss and gastrointestinal side effect, potentially unblinding participants making subjective endpoints more prone to bias.[16]

Possible solutions

Hospitalization is a key part of the natural history of HF and remains essential to capture in clinical trials. To improve the validity of trial outcomes, careful and deliberate trial designs are necessary to ensure the evidence generated is sufficiently robust. Possible solutions include:

1. Develop a standardized definition of HF hospitalization, which can be applied to make comparisons across clinical trial datasets.
2. Report the total hospitalization burden, not merely those for specific reasons, including recurrent episodes, and report the number needed to treat to avoid *any* hospitalization.
3. Use both investigator-reported and adjudicated outcomes, supported by routinely collected healthcare data, where possible.
4. In device trials, sham controls should be the new standard. We know these are feasible and ethically acceptable.
5. Where blinding is impossible, the default strategy should be to use less bias-prone harder endpoints such as death or total hospitalizations in adequately powered trials.
6. The use of probabilistic adjudication may yield valuable information from borderline events.

These are significant challenges which require physicians to look beyond the topline ‘positive’ results of mostly industry-sponsored trials. However, a better methodological approach is critical to avoid adopting costly new therapies that deliver only marginal or questionable benefit.

Graphical Abstract

Title: The changing landscape of heart failure trials.

Caption: Clinical trials which reduced all-cause or cardiovascular mortality are highlighted in dark green; those which did not reduce mortality but reduced total hospitalizations are highlighted in light green, those which did not reduce mortality and either did not reduce or did not report total hospitalizations are highlighted in amber.

*For the comparison of cardiac resynchronization therapy-defibrillator versus usual care. **For participants enrolled in the Americas.

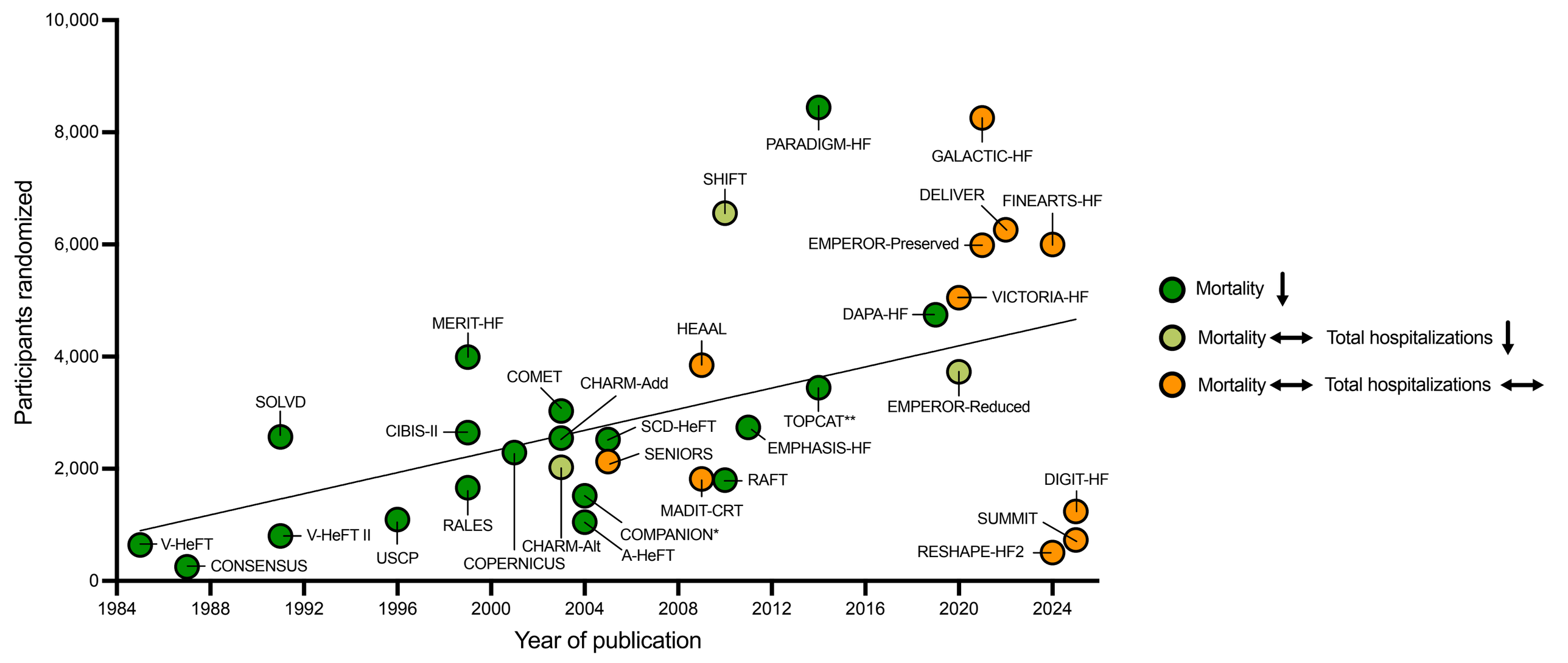
V-HeFT; Vasodilator-Heart Failure Trial, CONSENSUS; Cooperative North Scandinavian Enalapril Survival Study, SOLVD; Studies of Left Ventricular Dysfunction, USCP; US Carvedilol Heart Failure Study, RALES; Randomized Aldactone Evaluation Study, CIBIS; Cardiac Insufficiency Bisoprolol Study, COMET; Carvedilol or Metoprolol European Trial, MERIT-HF; Metoprolol CR/CL Randomized Intervention Trial in Congestive Heart Failure, COPERNICUS; Carvedilol Randomized Cumulative Survival, SCD-HeFT; Sudden Cardiac Death in Heart Failure Trial; CHARM; Candesartan in Heart Failure: Assessment of Reduction in Mortality and morbidity, CV; cardiovascular, HF; heart failure, COMPANION; Comparison of Medical Therapy, Pacing and Defibrillation in Heart Failure, CRT; cardiac resynchronization therapy, SENIORS; Study of the effects of nebivolol intervention on outcomes and rehospitalization in seniors with heart failure, A-HeFT; African-American Heart Failure Trial, QoL; quality of life, HEAAL; Heart failure Endpoint evaluation of Angiotensin II Antagonist Losartan, RAFT; Resynchronization-Defibrillation for Ambulatory Heart Failure Trial, MADIT-CRT; Multicentre Automatic Defibrillator Implantation Trial with Cardiac Resynchronization Therapy, SHIFT; Systolic Heart failure treatment with the If inhibitor Ivabradine Trial, EMPHASIS-HF;

Eplerenone in Mild Patients Hospitalization and Survival Study in Heart Failure, TOPCAT; Treatment of Preserved Cardiac Function Heart Failure with an Aldosterone Antagonist, PARADIGM-HF; Prospective comparison of ARNI with ACEi to Determine Impact on Global Mortality and morbidity in Heart Failure, VICTORIA-HF; Vericiguat Global Study in Subjects with Heart Failure with Reduced Ejection Fraction, DAPA-HF; Dapagliflozin and Prevention of Adverse Outcomes in Heart Failure, EMPORER-Reduced; Empagliflozin trial in Patients with chronic heart failure with reduced ejection fraction; GALACTIC-HF; Global Approach to Lowering Adverse Cardiac Outcomes through Improving Contractility in Heart Failure, EMPORER-Preserved; Empagliflozin trial in Patients with chronic heart failure with preserved ejection fraction, DELIVER; Dapagliflozin Evaluation to Improve the Lives of Patients with Preserved Ejection Fraction Heart Failure, FINEARTS-HF; Finerenone Trial to Investigate Efficacy and Safety Superior to Placebo in Patients with Heart Failure, DIGIT-HF; Digitioxin to Improve Outcomes in Patients with Advanced Chronic Heart Failure.

References

1. Packer M, Carver JR, Rodeheffer RJ, et al. Effect of Oral Milrinone on Mortality in Severe Chronic Heart Failure. *New England Journal of Medicine*. 1991;325(21):1468-1475. doi:10.1056/NEJM199111213252103
2. Cubbon RM, Gale CP, Kearney LC, et al. Changing Characteristics and Mode of Death Associated With Chronic Heart Failure Caused by Left Ventricular Systolic Dysfunction. *Circ Heart Fail*. 2011;4(4):396-403. doi:10.1161/CIRCHEARTFAILURE.110.959882
3. Solomon SD, McMurray JJV, Vaduganathan M, et al. Finerenone in Heart Failure with Mildly Reduced or Preserved Ejection Fraction. *New England Journal of Medicine*. 2024;391(16):1475-1485. doi:10.1056/NEJMoa2407107
4. The Effect of Digoxin on Mortality and Morbidity in Patients with Heart Failure. *New England Journal of Medicine*. 1997;336(8):525-533. doi:10.1056/NEJM199702203360801
5. Packer M. Development and Evolution of a Hierarchical Clinical Composite End Point for the Evaluation of Drugs and Devices for Acute and Chronic Heart Failure. *Circulation*. 2016;134(21):1664-1678. doi:10.1161/CIRCULATIONAHA.116.023538
6. Sorajja P, Whisenant B, Hamid N, et al. Transcatheter Repair for Patients with Tricuspid Regurgitation. *N Engl J Med*. 2023;388(20):1833-1842. doi:10.1056/NEJMoa2300525
7. Donal E, Dreyfus J, Leurent G, et al. Transcatheter Edge-to-Edge Repair for Severe Isolated Tricuspid Regurgitation: The Tri.Fr Randomized Clinical Trial. *JAMA*. 2025;333(2):124-132. doi:10.1001/jama.2024.21189
8. James S, Erlinge D, Storey RF, et al. Dapagliflozin in Myocardial Infarction without Diabetes or Heart Failure. *NEJM evidence*. 2024;3(2):EVIDoa2300286. doi:10.1056/EVIDoa2300286
9. Anker SD, Butler J, Filippatos G, et al. Empagliflozin in Heart Failure with a Preserved Ejection Fraction. *New England Journal of Medicine*. 2021;385(16):1451-1461. doi:10.1056/NEJMoa2107038
10. Mewton N, Donal E, Picard F, et al. Prognostic impact of precipitated cardiac decompensation in symptomatic heart failure with reduced ejection fraction and severe secondary mitral regurgitation. *Am Heart J*. 2023;265:83-91. doi:10.1016/j.ahj.2023.05.016
11. Carson P, Teerlink JR, Komajda M, et al. Comparison of Investigator-Reported and Centrally Adjudicated Heart Failure Outcomes in the EMPEROR-Reduced Trial. *JACC Heart Fail*. 2023;11(4):407-417. doi:10.1016/j.jchf.2022.11.017
12. Felker GM, Butler J, Januzzi JL, Desai AS, McMurray JJ V, Solomon SD. Probabilistic Readjudication of Heart Failure Hospitalization Events in the PARAGON-HF Study. *Circulation*. 2021;143(23):2316-2318. doi:10.1161/CIRCULATIONAHA.121.054496
13. Anker SD, Friede T, von Bardeleben RS, et al. Transcatheter Valve Repair in Heart Failure with Moderate to Severe Mitral Regurgitation. *N Engl J Med*. 2024;391(19):1799-1809. doi:10.1056/NEJMoa2314328
14. Généreux P, Schwartz A, Oldemeyer JB, et al. Transcatheter Aortic-Valve Replacement for Asymptomatic Severe Aortic Stenosis. *N Engl J Med*. 2025;392(3):217-227. doi:10.1056/NEJMoa2405880
15. Rajkumar CA, Nijjer SS, Cole GD, Al-Lamee R, Francis DP. “Faith Healing” and “Subtraction Anxiety” in Unblinded Trials of Procedures: Lessons from DEFER and

FAME-2 for End Points in the ISCHEMIA Trial. *Circ Cardiovasc Qual Outcomes*. 2018;11(3):e004665. doi:10.1161/CIRCOUTCOMES.118.004665
16. Packer M, Zile MR, Kramer CM, et al. Tirzepatide for Heart Failure with Preserved Ejection Fraction and Obesity. *New England Journal of Medicine*. 2025;392(5):427-437. doi:10.1056/NEJMoa2410027


Participants randomized
10,000
8,000
6,000
4,000
2,000
0
1984
1988
1992
1996
2000
2004
2008
2012
2016
2020
2024
Year of publication
V-HeFT
CONSENSUS
SOLVD
V-HeFT II
USCP
CIBIS-II
MERIT-HF
RALES
COPERNICUS
COMET
CHARM-Add
CHARM-Alt
SCD-HeFT
SENIORS
A-HeFT
COMPANION*
HEAAL
MADIT-CRT
RAFT
SHIFT
PARADIGM-HF
TOPCAT**
EMPHASIS-HF
GALACTIC-HF
DELIVER
FINEARTS-HF
EMPEROR-Preserved
EMPEROR-Reduced
DAPA-HF
VICTORIA-HF
DIGIT-HF
SUMMIT
RESHAPE-HF2
Mortality
Mortality
Total hospitalizations
Mortality
Total hospitalizations