JOURNAL OF THE AMERICAN COLLEGE OF CARDIOLOGY
© 2016 BY THE AMERICAN COLLEGE OF CARDIOLOGY FOUNDATION
PUBLISHED BY ELSEVIER

VOL. 68, NO. 24, 2016
ISSN 0735-1097/$36.00
http://dx.doi.org/10.1016/j.jacc.2016.09.936

REVIEW TOPIC OF THE WEEK

Natriuretic Peptides, 6-Min Walk Test, and Quality-of-Life Questionnaires as Clinically Meaningful Endpoints in HF Trials

João Pedro Ferreira, MD, PhD,[a] Kevin Duarte, Msc,[a] Todd L. Graves, PhD,[b] Michael R. Zile, MD,[c] William T. Abraham, MD,[d] Fred A. Weaver, MD,[e] JoAnn Lindenfeld, MD,[f] Faiez Zannad, MD, PhD[a]

ABSTRACT

The Expedited Access for Premarket Approval and De Novo Medical Devices Intended for Unmet Medical Need for Life Threatening or Irreversibly Debilitating Diseases or Conditions document was issued as a guidance for industry and for the Food and Drug Administration. The Expedited Access Pathway was designed as a new program for medical devices that demonstrated the potential to address unmet medical needs for life threatening or irreversibly debilitating conditions. The Food and Drug Administration would consider assessments of a device's effect on intermediate endpoints that, when improving in a congruent fashion, are reasonably likely to predict clinical benefit. The purpose of this review is to provide evidence to support the use of 3 such intermediate endpoints: natriuretic peptides, such as N-terminal pro–B-type natriuretic peptide/B-type natriuretic peptide, the 6-min walk test distance, and health-related quality of life in heart failure.
(J Am Coll Cardiol 2016;68:2690–707) © 2016 by the American College of Cardiology Foundation.

Large-scale event-driven heart failure (HF) trials are increasingly more challenging to conduct. Demonstrating a significant benefit for morbidity and mortality rate reduction on top of the current therapy has become increasingly difficult, requiring larger and longer trials that have become prohibitively expensive (1). This is particularly true for the case of device trials, which pose additional difficulties with patient enrollment, costs, and device iteration.

On April 13, 2015, the Expedited Access for Premarket Approval and De Novo Medical Devices Intended for Unmet Medical Need for Life Threatening or Irreversibly Debilitating Diseases or Conditions document was issued as guidance for industry and Food and Drug Administration (FDA) Center for Devices and Radiological Health staff (2). The Expedited Access Pathway (EAP) was designed as a new program for medical devices to address unmet medical needs for life-threatening or irreversibly debilitating conditions. The FDA Center for Devices and Radiological Health intended the EAP program to, "help patients have more timely access to medical devices by expediting their development, assessment and review, while preserving the FDA's statutory

Listen to this manuscript's audio summary by *JACC* Editor-in-Chief Dr. Valentin Fuster.

From [a]INSERM, Centre d'Investigations Cliniques Plurithématique 1433, Université de Lorraine, CHRU de Nancy and F-CRIN INI-CRCT, Nancy, France; [b]Berry Consultants LLC, Austin, Texas; [c]Medical University of South Carolina and the RHJ Department of Veterans Affairs Medical Center, Charleston, South Carolina; [d]Ohio State University, Columbus, Ohio; [e]Division of Vascular Surgery and Endovascular Therapy, Department of Surgery, University of Southern California, Los Angeles, California; and the [f]Vanderbilt Heart and Vascular Institute, Nashville, Tennessee. Dr. Ferreira has received board membership fees from Novartis. Dr. Zile is a consultant to and receives research support from CVRx, Novartis, Abbott, and Medtronic. Dr. Abraham has received consulting fees and speaking honoraria from CVRx; and is member of the CVRx Heart Failure Executive Steering Committee. Dr. Weaver is a consultant to and receives research support from CVRx, WL Gore, and Endologix. Dr. Lindenfeld is a consultant to Novartis, CVRx, Abbott, St. Jude, Resmed, and Relyspsa; and receives research support from Novartis. Dr. Zannad is consultant to Boston Scientific, CVRx, AstraZeneca, Boehringer, Amgen, GE Healthcare, Quantum Genomics, Actelion, Bayer, Janssen, Resmed, Roche Diagnostics, Relypsa, ZS Pharma, Takeda, and Novartis. All other authors have reported that they have no relationships relevant to the contents of this paper to disclose.

Manuscript received July 18, 2016; revised manuscript received September 6, 2016, accepted September 19, 2016.

standard for premarket approval." This guidance document also described the types of clinical evidence that might be used to support approval of a device under the EAP and, in particular, discussed the use of intermediate and surrogate endpoints.

For clarity, it is important to define the terms intermediate and surrogate endpoints. An intermediate endpoint is used in a clinical study as a measure of clinical benefit or risk concerning a symptom or measure of function that is not the ultimate outcome of the disease (e.g., exercise tolerance and quality of life [QoL] in trials of device treatments for HF). In contradistinction, a surrogate endpoint is not itself a measure of clinical benefit, but is used in trials as a substitute that is reasonably likely to predict clinical benefit, on the basis of epidemiological, therapeutic, pathophysiological, or other sources of scientific evidence. The types of measurements that may be used as surrogate endpoints are laboratory or medical imaging measurements, or physical signs (e.g., natriuretic peptides [NPs] in HF trials, as a surrogate for HF hospitalization and cardiovascular [CV] death clinical endpoints).

The FDA would consider assessments of a device's effect on intermediate and surrogate endpoints that, when improving in a congruent fashion, are "reasonably likely to predict clinical benefit." For the purpose of this paper, we will analyze NPs, such as N-terminal-pro-B-type natriuretic peptide (NT-proBNP)/B-type natriuretic peptide (BNP), the 6-min walk distance test (6MWD), and health-related QoL (3,4), which are referred to as clinically meaningful endpoints because they may be used for the assessment of clinical benefit or risk at the patient level.

The proposal to use clinically meaningful endpoints should include evidence that is reasonably likely to predict clinical benefit that is generally of value to patients, "even if this does not indisputably lead to reduced morbidity or mortality," enhancing the probability of approval for a novel therapy when no alternative treatment is available, or provide the basis for potential effectiveness in a phase III outcome trial. The use of clinically meaningful endpoints in clinical drug and device development may allow for pretesting the efficacy and safety of a treatment in phase 2 and proof-of-concept trials, as well as in pre-market access trials of devices.

The purpose of this review is to provide evidence to support the use of NPs, 6MWD, and QoL as clinically meaningful endpoints. Both the clinical benefit evidenced by changes in these endpoints and the likelihood that these changes predict morbidity and mortality will be examined. We also try to identify knowledge gaps and to suggest a potential standard application for these intermediate endpoints in clinical drug and device development programs.

NATRIURETIC PEPTIDES

NPs are cardiac hormones produced by cardiac myocytes in response to changes in myocardial wall stress. Plasma concentrations of NPs have been shown to be useful for the diagnosis and management of HF patients (5,6). For example, NP levels have been correlated with left ventricular end-diastolic pressure, New York Heart Association (NYHA) functional class, exercise tolerance, and pulmonary capillary wedge pressure (7–9). More importantly, NP levels have been associated with clinical event rates, using either baseline levels or changes between levels at baseline and at a given time point (10–12).

HEART FAILURE WITH REDUCED EJECTION FRACTION

The association between changes in NP levels and treatment effect has been most extensively studied in patients with HF with reduced ejection fraction (HF-REF) (4).

A meta-analysis studied the effects of treatment on BNP (n = 15 trials) and NT-proBNP (n = 6 trials). This analysis identified the change in BNP and NT-proBNP from each individual trial to assess the association between NP change and treatment effect. The odds ratio for all-cause mortality was not correlated with the drug/device-induced BNP change (Spearman's rho = -0.65, p = 0.818) or NT-proBNP (Spearman's rho = -0.67, p = 0.148). Given these results, the investigators suggest that therapy-induced changes in NPs do not correlate with changes in mortality (4). However, this meta-analysis did not incorporate recent HF trials, presented greatly different lengths of follow-up and NP measurement time intervals, and did not use longitudinal individual patient data. In addition, there was great variation in the range of treatment effects on NP levels between studies. For example, the trial that contributed the most patient data in this meta-analysis was the COPERNICUS (Effect of carvedilol on the morbidity of patients with severe chronic heart failure: results of the carvedilol prospective randomized cumulative survival) study (13), with 1,048 patients. In this study, carvedilol reduced all-cause mortality by ~35% as compared with placebo; however, no significant differences in median NT-proBNP levels between the treatment and

ABBREVIATIONS AND ACRONYMS

6MWD = 6-min walk test distance

AHF = acute heart failure

BNP = B-type natriuretic peptide

CV = cardiovascular

HF = heart failure

HF-REF = heart failure with reduced ejection fraction

NP = natriuretic peptide

NYHA = New York Heart Association

NT-proBNP = N-terminal pro-B-type natriuretic peptide

QoL = quality of life

placebo groups were found during follow-up. However, when analyzing the proportion of change from baseline, carvedilol induced a more pronounced decrease in NT-proBNP levels at 4 and 7 months (-17% in carvedilol vs. -5% in placebo at 4 months, and -28% in carvedilol vs. -5% in placebo at 7 months; $p < 0.01$ for both 4 and 7 months). This finding may be explained by a greater reduction of outliers (i.e., extremely high values) by carvedilol treatment as compared with placebo (but unfortunately these data were not shown in the paper) (13,14). From these data, we may conclude that carvedilol induced NT-proBNP changes that were congruent with its effects on morbidity and mortality. A more recent meta-analysis (15), analyzed the association between the changes in NP levels and treatment effect on HF hospitalizations in HF-REF. This study found that active treatments significantly reduced the risk of HF hospitalization, and that changes in BNP and NT-proBNP were significantly associated with reductions in HF hospitalizations, suggesting that NPs can be used as surrogate markers for HF hospitalization in HF-REF.

Many other relevant examples of the association between NP changes, treatment effects, and outcomes exist. In the CARE-HF (Effect of Cardiac Resynchronization on Morbidity and Mortality in Heart Failure) trial (16), patients in the cardiac resynchronization therapy (CRT) arm had significantly lower median NT-proBNP at 3 and 18 months, compared with baseline (1,920 [interquartile range (IQR): 744 to 4,288] pg/ml vs. 1,809 [IQR: 719 to 3,949] pg/ml at baseline; -537 pg/ml in median NT-proBNP in favor of CRT therapy at 3 months, $p < 0.001$; and -567 pg/ml in favor of CRT at 18 months, $p < 0.001$) (17). Overall CRT was associated with lower all-cause death rates (hazard ratio [HR]: 0.64; 95% confidence interval [CI]: 0.48 to 0.85; $p < 0.002$). These data suggest an association between mortality reduction and lower NP levels. In the AREA IN-CHF (Anti-remodelling effect of canrenone in patients with mild chronic heart failure) trial, canrenone produced a significant decrease in BNP levels from baseline to 6 months (-37% in the canrenone arm vs. -8% in the placebo arm; $p < 0.001$). The composite endpoint of cardiac death and HF hospitalization was significantly lower in the canrenone arm at 12 months (7.9% in the canrenone arm vs. 15.1% in the placebo arm; $p = 0.02$) (18). In the PARADIGM-HF (Angiotensin-Neprilysin Inhibition versus Enalapril in Heart Failure) trial, patients taking LCZ696 had significantly lower median NT-proBNP at 4 weeks and at 8 months (19). The trial was stopped at 27 months due to LCZ696 superiority (HR: 0.80; 95% CI: 0.73 to 0.87; $p < 0.001$) for the primary outcome of hospitalization for HF or CV death. In the ASTRONAUT (Effect of Aliskiren on Postdischarge Mortality and Heart Failure Readmissions Among Patients Hospitalized for Heart Failure) trial, aliskiren was associated with a statistically significant decrease in the adjusted geometric mean NT-proBNP level compared with placebo at each time point tested (months 1, 6, and 12), but p values for other methods used to estimate BNP variation (e.g., delta or absolute differences) have not been provided. However, the median values between the treatment and placebo groups looked similar. For example, after 6 months of trial, the median NT-proBNP was 1,754 pg/ml (IQR: 864 to 3,641 pg/ml) in the aliskiren group versus 1,897 pg/ml (IQR: 1,015 to 3,781 pg/ml) in placebo group. This was likely to be consistent with the neutral effect of aliskiren on CV death or HF rehospitalization at 6 months (HR: 0.92; 95% CI: 0.76 to 1.12 at 6 months; $p = 0.41$) (20). These potential differences between mathematical formulas will be discussed later in the present paper.

The efficacy of baroreflex activation therapy (BAT) in advanced HF was examined in a robust feasibility study (the HOPE4HF [Hope for Heart] study) (21). The 3 intermediate endpoints (NT-proBNP, 6MWD, QoL) that constitute the focus of this review were examined and compared in control subjects (treated with guideline-directed medical therapy) versus those with guideline-directed medical therapy plus BAT. NT-proBNP was reduced in the BAT treatment group and increased in the control group, with a significant between-group difference (median -69.0 pg/ml [IQR: -504 to 198 pg/ml] vs. 129.5 pg/ml [IQR: -67 to 619 pg/ml]; $p = 0.02$). Although exploratory, in the BAT group there was a significant reduction in the rate of HF hospitalization from pre- to post-enrollment (0.63 ± 1.5 hospitalizations/patient/year to 0.14 ± 0.5 hospitalizations/patient/year; $p = 0.01$), with no change seen in the control group (0.36 ± 1.1 to 0.31 ± 0.97 hospitalizations/patient/year; $p = 0.85$). However, the between-group difference in the post-randomization rate of HF hospitalization did not reach statistical significance ($p = 0.35$). In addition, BAT was associated with a trend toward fewer days hospitalized for HF ($p = 0.08$). BAT will be further examined in the Beat-HF trial (NCT02627196), a pivotal phase III trial, using the EAP.

These and other examples of pharmacological and device trials with NP endpoints are provided in **Table 1**. In general, when a HF-REF trial is clearly positive for a clinical benefit on morbidity and mortality, NP levels fall significantly.

TABLE 1 Evidence for the Use of Surrogate Markers as Endpoints in Chronic and Symptomatic HF-REF

First Author (Year; Trial Name) (Ref. #)	Surrogate Marker Change	Outcome	Comments/Agreement Between Test and Outcomes
BNP/NT-proBNP			
McKelvie et al. (1999; RESOLVD pilot study) (77)	Candesartan + enalapril reduced delta log BNP compared with candesartan and enalapril alone	Lower increment in EDV and ESV with candesartan + enalapril as compared with candesartan and enalapril alone	DB-RCT. Agreement: yes
Conraads et al. (2004) (78)	Combined endurance/resistance training reduced mean NT-proBNP levels at 4 months	Reduction in mean NT-proBNP was the primary outcome	Nonrandomized parallel-group study with 27 patients. The primary outcome was NT-proBNP
Hartmann et al. (2004; data from the COPERNICUS trial) (13,14)	Median NT-proBNP values between carvedilol and placebo groups were not significantly different at baseline, 1 month, 4 months, and study end. However, compared with placebo, carvedilol induced a steeper decrease in NT-proBNP levels (in % of change relative to baseline) at 4 and 7 months: -17% vs. -5% and -28% vs. -5%, respectively	Carvedilol reduced mortality compared with placebo (HR: 0.65; 95% CI: 0.52–0.81)	Data derived from DB-RCT. The association of NT-proBNP changes and outcome were not tested. The steeper decrease in % NT-proBNP relative to baseline in the carvedilol group (but no difference in median values) may be explained by a greater reduction of extremely high values by carvedilol treatment, but these data are not shown. Agreement: yes (for absolute delta changes)
Krum et al. (2007) (79)	Rosuvastatin did not reduce mean or % change in BNP levels at 6 months compared to placebo	Rosuvastatin did not affect the primary endpoint of change in LVEF or any of the other pre-specified endpoints	DB-RCT. Agreement: yes
Toblli et al. (2007) (50)	Iron therapy reduced mean NT-proBNP levels compared with placebo at 6 months	Iron therapy improved functional capacity, symptoms, and QoL	DB-RCT. Agreement: yes
Fruhwald et al. (2007; from CARE-HF data) (17)	Patients in CRT arm had significantly lower median NT-proBNP at 3 and 18 months (but not at baseline) At baseline, the median plasma concentration of NT-proBNP was similar in patients assigned to CRT or medical therapy, 1,920 (IQR: 744–4,288) pg/ml vs. 1,809 (IQR: 719–3,949) pg/ml. The differences in medians between the CRT and medical therapy groups were highly significant at both 3 months (537 pg/ml; $p < 0.001$) and 18 months of follow-up (567 pg/ml; $p < 0.001$)	CRT is associated with lower death rates (HR: 0.64; 95% CI: 0.48 to 0.85; $p < 0.002$)	DB-RCT. CRT exerts an early and sustained reduction in NT-proBNP, reflecting improvements in ventricular function. Agreement: yes
Masson et al. (2008; from Val-HeFT data) (3)	Steady levels or reduction (in % quartiles) in NT-proBNP from baseline to 4 months were associated with lower mortality rates as compared with the highest quartile (>33% NT-proBNP increase)	Death rates of 9.2% in quartile 1 (reduction >31% at 4-month NT-proBNP relative to baseline) to 21.5% in quartile 4 (increase >33%)	Data derived from DB-RCT. Absolute changes in NT-proBNP were not consistently associated, possibly because patients with highest baseline NT-proBNP (associated with dismal outcomes) had also great reductions. Observational studies limitations. Agreement: NA
Pascual-Figal et al. (2008) (80)	≥26% (median) reduction at 2 weeks after decompensation was associated with ~23% lower 1-year event rates. In multivariable Cox models, each 10% NT-proBNP reduction was associated with ~17% fewer event rates at 1 year	Lower event rates at 1 year in patients with NT-proBNP reduction	Observational study in 71 patients. Any absence of decline associated with worse outcomes. Agreement: NA
Boccanelli et al. (2009; AREA IN-CHF trial) (18)	Canrenone produced a steeper decrease in delta BNP levels at 6 months (-37% in the canrenone arm vs. -8% in the placebo arm)	The composite endpoint of cardiac death and hospitalization was significantly lower in the canrenone arm at 12 months (7.9% in canrenone arm vs. 15.1% in placebo arm; $p = 0.02$)	DB-RCT. Agreement: yes
Bielecka-Dabrowa et al. (2009) (81)	Atorvastatin therapy reduced mean NT-proBNP levels at 2 months, whereas NT-proBNP levels increased in the control group	Atorvastatin reduced hospitalizations due to HF	Randomized open-label in 68 patients with DCM. Agreement: yes
Solomon et al. (2012; PARAMOUNT trial) (25)	LCZ696 significantly reduced NT-proBNP at 12 weeks as compared with valsartan (LCZ696 baseline = 783 [IQR: 670–914] pg/ml, LCZ696 12 weeks = 605 [IQR: 512–714] pg/ml; valsartan baseline = 862 [IQR: 733–1,012] pg/ml, valsartan 12 weeks = 835 [IQR: 710–981] pg/ml; ratio LCZ696/valsartan = 0.77 [95% CI: 0.64–0.92], $p = 0.005$)	Phase II trial. No CV outcomes measured.	DB-RCT. Agreement: NA

Continued on the next page

TABLE 1 Continued

First Author (Year; Trial Name) (Ref. #)	Surrogate Marker Change	Outcome	Comments/Agreement Between Test and Outcomes
Gheorghiade et al. (2013; ASTRONAUT trial) (20)	Aliskiren was associated with a statistically significant decrease in the adjusted geometric mean NT-proBNP level compared with placebo at each time point tested (months 1, 6, and 12). Randomization: 2,838 (IQR: 1,516–5,235) pg/ml in aliskiren vs. 2,674 (IQR: 1,552–5,234) pg/ml in placebo 1 month: median NT-proBNP = 2,433 (IQR: 1,275–4,387) pg/ml in aliskiren vs. 2,522 (IQR: 1,455–4,625) pg/ml in placebo Adjusted geometric mean NT-proBNP = 0.86 (0.81–0.91) in aliskiren vs. 0.95 (0.90–1.00) in placebo Ratio aliskiren/placebo = 0.90 (0.84–0.97); p = 0.01 6 months: median NT-proBNP = 1,754 (IQR: 864–3,641) pg/ml in aliskiren vs. 1,897 (IQR: 1,015–3,781) pg/ml in placebo Adjusted geometric mean NT-proBNP = 0.64 (0.59–0.70) in aliskiren vs. 0.76 (0.70–0.82) in placebo Ratio aliskiren/placebo = 0.85 (0.76–0.95); p < 0.01 12 months: median NT-proBNP = 1,576 (IQR: 681–3,156) pg/ml in aliskiren vs. 1,792 (IQR: 887–3,518) pg/ml in placebo Adjusted geometric mean NT-proBNP = 0.62 (0.57–0.69) in aliskiren vs. 0.74 (0.67–0.82) in placebo Ratio aliskiren/placebo = 0.84 (0.73–0.96); p < 0.01	Aliskiren did not reduce CV death or HF rehospitalization at 6 months (HR: 0.92; 95% CI: 0.76–1.12; p = 0.41) or 12 months after discharge (HR: 0.93; 95% CI: 0.79–1.09; p = 0.36)	DB-RCT. Agreement: no
Zannad et al. (2014; NECTAR-HF trial) (82)	No significant differences at 6 months in median NT-proBNP between therapy vs. controls	VNS failed to demonstrate a significant effect on primary and secondary endpoint measures of cardiac remodeling and functional capacity in symptomatic HF patients	DB-RCT. Agreement: yes
Packer et al. (2015; from PARADIGM-HF data) (19)	NT-proBNP was similar between LCZ969 and enalapril groups at baseline, ~1,300 pg/ml. Patients in the LCZ696 group had lower median NT-proBNP at 4 weeks ~800 (IQR: 600–1,500) pg/ml vs. 1,250 (750–2,100) pg/ml in the enalapril group, p < 0.001 and at 8 months ~800 (IQR: 400–1,600) pg/ml in the LCZ969 group vs. 1,250 (IQR: 700–2,100) pg/ml in the enalapril group, p < 0.001	The trial was stopped at 27 months due to LCZ696 superiority (HR: 0.80; 95% CI: 0.73–0.87; p < 0.001) for the primary outcome of hospitalization for HF or CV death	DB-RCT. LCZ696 exerts an early and sustained reduction in NT-proBNP. Agreement: yes
Ordu et al. (2015) (83)	Patients in the ivabradine group had lower mean NT-proBNP at 6 months compared with the control group. NT-proBNP at baseline = 1,353 ± 1,453 pg/ml in the ivabradine group vs. 1,383 ± 1,064 pg/ml in the control group NT-proBNP at 6 months = 717 ± 834 pg/ml in the ivabradine group vs. 1,323 ± 979 pg/ml in the control group	In the SHIFT trial, after a median follow-up of ~23 months, ivabradine reduced the primary outcome of hospitalization for HF or CV death (HR: 0.82; 95% CI: 0.75–0.90; p < 0.001)	Open-label parallel-group study with 98 patients. Agreement: yes
Gheorghiade et al. (2015; SOCRATES-REDUCED trial) (84)	Pooled vericiguat doses showed no significant difference in 12-week log-transformed NT-proBNP levels. However, patients in the highest vericiguat dose groups (2.5 to 10 mg) had significant reductions in the ratios of geometric mean changes from baseline to week 12. Pooled vericiguat group Log-transformed: baseline = 7.969; 12 weeks = 7.567; difference = −0.402; Geometric means: baseline = 2,890 pg/ml; 12 weeks = 1,932 pg/ml Placebo Log-transformed: baseline = 8.283; 12 weeks = 8.002; difference = −0.280; Geometric means: baseline = 3,955 pg/ml; 12 weeks = 2,988 pg/ml Difference of means = −0.122 (−0.32 to 0.07) Ratio of geometric means, 0.885 (0.73–1.08); p = 0.15	Change from baseline to week 12 in log-transformed NT-proBNP level was the primary endpoint. Phase 3 trial underway	DB-RCT. Agreement: NA
Mebazaa et al. (2007; from the SURVIVE trial) (28)	Levosimendan group had greater decreases in BNP level as compared with dopamine: Mean change of around -631 pg/ml in the levosimendan group vs. -397 pg/ml in the dobutamine group; p < 0.001. Overlapping changes persisted through 5 days compared with the dobutamine group (p < 0.001 for all time points)	No statistical differences were observed between treatment groups for the primary outcome of all-cause mortality at 180 days (HR: 0.91; 95% CI: 0.74–1.13; p = 0.40) or any other secondary endpoints (all-cause mortality at 31 days, number of days alive and out of the hospital, patient global assessment, patient assessment of dyspnea at 24 h, and CV mortality at 180 days)	DB-RCT. Agreement: no

Continued on the next page

TABLE 1 Continued

First Author (Year; Trial Name) (Ref. #)	Surrogate Marker Change	Outcome	Comments/Agreement Between Test and Outcomes
Metra et al. (2013; from RELAX-AHF trial) (31)	Serelaxin group had significant decrease in NT-proBNP levels from baseline to day 2 NT-proBNP: Baseline geometric mean = 5,003.50 in placebo vs. 5,125.46 in serelaxin Day 2 geometric mean = 3,037.50 in placebo vs. 2,544.23 in serelaxin Geometric mean change = 0.607 in placebo vs. 0.492 in serelaxin; $p < 0.001$ 30% decrease from baseline to day 2 = 58.0% in placebo vs. 69.0% in serelaxin; $p < 0.001$	Serelaxin improved the VAS AUC primary dyspnea endpoint compared with placebo, but had no significant effect on the other primary endpoint Likert scale. No significant effects were recorded for the secondary endpoints of cardiovascular death or readmission to hospital for HF or renal failure until day 180 (HR = 1.02; 95% CI: 0.74–1.41; $p = 0.89$) or days alive out of the hospital up to day 60. Fewer deaths at day 180 were observed in the serelaxin group (HR 0.63; 95% CI: 0.42–0.93; $p = 0.019$)	DB-RCT. Agreement: not for the coprimary nor secondary endpoints.
6MWT			
Guyatt et al. (1985) (36)	6MWT was reproducible, and correlated with the conventional measures of functional status and exercise capacity.	Validation study.	Observational study. Agreement: NA
Lipkin et al. (1986) (85)	Patients with HF-REF on NYHA functional class III walked less than those on NYHA functional class II (402 m vs. 558 m; $p < 0.01$)	Group comparison.	Observational study. Distance walked varied little in patients with a high maximal oxygen consumption. No delta differences are provided. Agreement: NA
Roul et al. (1998) (46)	Patients who walked ≤300 m had a worse prognosis compared with those walking farther	Group comparison	Observational study. No delta differences are provided. Agreement: NA
McKelvie et al. (1999; RESOLVD pilot study) (77)	During the 43 weeks of the study, there were no significant differences in 6MWT between groups (candesartan + enalapril vs. candesartan alone vs. enalapril alone)	Lower increment in EDV and ESV with candesartan + enalapril compared with candesartan and enalapril alone	DB-RCT. Agreement: no
Cazeau et al. (2001) (42)	The mean distance in the 6MWT was ~22% greater after 3 months with active pacing (399 ± 100 m vs. 326 ± 134 m, $p < 0.01$)	QoL and peak oxygen consumption also improved and hospitalizations were reduced	Single-blind, crossover, RCT. Agreement: yes
Abraham et al. (2002) (43)	Patients assigned to CRT experienced an improvement of ~26% in the 6MWT (+39 m vs. +10 m, $p < 0.01$)	QoL, time on the treadmill, and LVEF also improved	Single-blind RCT. Agreement: yes
Auricchio et al. (2003) (49)	CRT in patients with QRS interval >150 ms improved 6MWT in ~10% (407 ± 61 to 447 ± 72 m), whereas patients with thin QRS intervals did not	CRT significantly improved the maximal and submaximal exercise capacity, QoL, and functional status of HF patients, especially those with QRS intervals >150 ms	Single-blind, crossover, RCT. Agreement: yes
Higgins et al. (2003) (39)	CRT improved 6MWT from baseline to 6 months compared with the no CRT group (35 ± 7 m vs. 15 ± 7 m, $p < 0.05$)	CRT improved functional status in patients indicated for an ICD who also have symptomatic HF and intraventricular conduction delay	Single-blind, crossover, RCT. Agreement: yes
Young et al. (2003; MIRACLE ICD trial) (41)	Patients assigned to CRT presented no differences in the change of 6MWT compared to control (55 m vs. 53 m; $p = 0.36$)	CRT improved QoL, functional status, and exercise capacity in patients with moderate to severe HF, a wide QRS interval, and life-threatening arrhythmias	DB-RCT. Agreement: no
Abraham et al. (2004; from MIRACLE ICD II trial) (38)	CRT did not improve 6MWT as compared to ICD in NYHA functional class II patients with a QRS interval >130 ms	CRT improved cardiac structure and function, and composite clinical response over 6 months	DB-RCT. Agreement: no
Toblli et al. (2007) (50)	Iron therapy improved mean 6MWT from baseline to 6 months between iron therapy and placebo groups (184.5 ± 58.5 m vs. 240.1 ± 51.2 m; $p < 0.05$)	Iron therapy improved symptoms, QoL and reduced NT-proBNP	DB-RCT. Delta improvement in meters was not calculated, although differences were statistically significant. Agreement: yes
Guazzi et al. (2009) (86)	No significant differences were observed in 6MWT distance between patients who remained alive vs. those who died at 4 years.	6MWT was not significantly associated with the outcome of death	Observational study. No delta comparisons where provided. CPET-derived variables were associated with the outcome. Agreement: NA

Continued on the next page

TABLE 1 Continued

First Author (Year; Trial Name) (Ref. #)	Surrogate Marker Change	Outcome	Comments/Agreement Between Test and Outcomes
Täger et al. (2014) (87)	Submaximal exercise capacity as represented by the 6MWT varies little in stable HF-REF patients for up to 1-year intervals. The MID for changes in 6MWT values over a period of 6 to 12 months is ~36 m.	Group comparison and validation study	Observational study. One SE of measurement was used as a proxy for MID in 6MWT. Agreement: NA
Ingle et al. (2014) (88)	Distance in 6MWT is an independent predictor of mortality in patients with HF-REF both at baseline and at 1 year	All-cause death: ~3% mortality reduction at 8 years for each meter increment in 6MWT, both at baseline and at 1 year	Observational study. Delta values were not provided. Agreement: NA
Gold et al. (2016; from INOVATE-HF trial) (44)	VNS improved 6MWT distance (+28.2 m in the VNS group vs. -4.6 m in the control group; $p < 0.01$)	VNS did not improve the primary efficacy outcome of death or HF hospitalization (HR: 1.14; 95% CI: 0.86–1.53; $p = 0.37$)	Open RCT Agreement: No
QoL			
McKelvie et al. (1999; RESOLVD pilot study) (77)	Candesartan + enalapril did not improve QoL scores as compared to candesartan and enalapril alone	Lower increment in EDV and ESV with candesartan + enalapril as compared with candesartan and enalapril alone	DB-RCT. Agreement: no
Cazeau et al. (2001) (42)	Active pacing improved QoL by ~32%, as assessed by the MLHFQ	6MWT and peak oxygen consumption also improved and hospitalizations were reduced	Single-blind, crossover, RCT. Agreement: yes
Abraham et al. (2002) (43)	Patients assigned to CRT experienced improvement in QoL (-18 points vs. -9 points, $p < 0.01$), as assessed by the MLHFQ	6MWT, time in the treadmill, and LVEF also improved	Single-blind RCT Agreement: yes
Auricchio et al. (2003) (49)	CRT in patients with QRS interval >150 ms improved QoL, as assessed by the MLHFQ (17 ± 13 points vs. 25 ± 15 points, $p < 0.01$)	CRT significantly improved the maximal and submaximal exercise capacity, 6MWT, and functional status of HF patients, especially in patients with QRS intervals >150 ms	Single-blind, crossover, RCT Agreement: yes
Higgins et al. (2003) (39)	CRT improved QoL from baseline to 6 months compared with the no CRT group, as assessed by the MLHFQ (-17.5 vs. -11.0, $p = 0.02$)	CRT improved functional status in patients indicated for an ICD who also have symptomatic HF and intraventricular conduction delay	Single-blind, crossover, RCT Agreement: yes
Abraham et al. (2004; from MIRACLE ICD II trial) (38)	CRT did not improve QoL scores as compared with ICD in NYHA functional class II patients with a QRS interval >130 ms	CRT improved cardiac structure and function and composite clinical response over 6 months	DB-RCT Agreement: no
Tate et al. (2007; from BEST trial) (51)	QoL questionnaires were strongly associated with mortality. At 12 months, bucindolol-treated patients had improvement according to some QoL questionnaires, but not according to others.	Beta-blocker therapy is mainstay treatment in HF-REF	DB-RCT Agreement: yes
Toblli et al. (2007) (50)	Iron therapy improved QoL compared with placebo at 6 months, as assessed by the MLHFQ (59 ± 8 points vs. 41 ± 7 points; $p < 0.05$)	Iron therapy improved functional capacity, symptoms, and NT-proBNP	DB-RCT Agreement: yes
Zannad et al. (2014; NECTAR-HF trial) (82)	VNS improved QoL, as assessed by the MLHFQ, at 6 months compared with controls (35.8 ± 20.8 points vs. 41.8 ± 24.3 points; $p < 0.05$)	VNS failed to demonstrate a significant effect on primary and secondary endpoint measures of cardiac remodeling and functional capacity in symptomatic HF patients	DB-RCT Agreement: no
O'Connor et al. (53), and Flynn et al. (52) (2009; ACTION-HF)	Sustained improvements in KCCQ in exercise arm	Primary endpoint of all-cause mortality or hospitalization was negative in primary protocol-specified analysis	RCT Agreement: no
Anker et al. (2009; FAIR-HF) (54)	Improvements in QoL KCCQ at 4, 12, and 24 weeks	Patient-reported global assessment and NYHA functional class both improved at 24 weeks	RCT Agreement: yes
Costanzo et al. (2012; PEERLESS-HF) (56)	Improvement in KCCQ in the device arm	Study stopped after enrollment of 122 patients for futility on the primary endpoint of VO_2	RCT Agreement: no
Moss et al. (2009) (57) and Veazie et al. (2012) (58) (MADIT-CRT)	Improvements in KCCQ in LBBB group only	Primary endpoint of all-cause death and HF events showed significant reduction in the CRT arm	RCT Agreement: mixed Overall positive study, but QoL analysis positive only in LBBB in subgroup
Ponikowski et al. (2015; CONFIRM-HF) (55)	Significant improvement in KCCQ at 24 weeks	Significant improvement with the primary outcome of 6MWT at 24 weeks	RCT Agreement: yes

Continued on the next page

TABLE 1 Continued

First Author (Year; Trial Name) (Ref. #)	Surrogate Marker Change	Outcome	Comments/Agreement Between Test and Outcomes
Pitt et al. (2014) (59) and Lewis et al. (2016) (60) (TOPCAT)	Spironolactone improved KCCQ at 4, 12, and 36 months	Composite of death from CV causes, aborted cardiac arrest, or hospitalization for the management of HF not improved	DB-RCT Agreement: no
Gold et al. (2016; from INOVATE-HF trial) (44)	VNS improved QoL, as assessed by the KCCQ score (+11.2 points in the VNS group vs. +6.9 points in the control group; $p < 0.01$).	VNS did not improve the primary efficacy outcome of death or HF hospitalization (HR: 1.14; 95% CI: 0.86–1.53; $p = 0.37$)	Open RCT Agreement: No

6MWT = 6-min walking test; ACTION-HF = Efficacy and Safety of Exercise Training in Patients With Chronic Heart Failure; AREA IN-CHF = Anti-remodelling effect of canrenone in patients with mild chronic heart failure trial; ASTRONAUT = Effect of Aliskiren on Postdischarge Mortality and Heart Failure Readmissions Among Patients Hospitalized for Heart Failure; BEST = Beta-Blocker Bucindolol in Patients with Advanced Chronic Heart Failure; BNP = B-type natriuretic peptide; CARE-HF = The Effect of Cardiac Resynchronization on Morbidity and Mortality in Heart Failure trial; CI = confidence interval; CONFIRM-HF = Beneficial effects of long-term intravenous iron therapy with ferric carboxymaltose in patients with symptomatic heart failure and iron deficiency; COPERNICUS = Carvedilol prospective randomized cumulative survival trial; CRT = cardiac resynchronization therapy; CV = cardiovascular; DB-RCT = double-blind randomized controlled trial; DCM = dilated cardiomyopathy; EDV = end-diastolic volume; ESV = end-systolic volume; HF = heart failure; HF-REF = heart failure with reduced ejection fraction; HR = hazard ratio; ICD = implantable cardioverter-defibrillator; INOVATE-HF = Vagus Nerve Stimulation for the Treatment of Heart Failure; IQR = interquartile range; KCCQ = Kansas City Cardiomyopathy Questionnaire; LBBB = left bundle branch block; LVEF = left ventricular ejection fraction; MID = minimal important difference; MIRACLE ICD = Combined cardiac resynchronization and implantable cardioversion defibrillation in advanced chronic heart failure; MLHFQ = Minnesota Living with Heart Failure Questionnaire; NECTAR-HF = NEural Cardiac TherApy foR Heart Failure; NT-proBNP = N-terminal pro–B-type natriuretic peptide; NYHA = New York Heart Association; PARADIGM-HF = Angiotensin-Neprilysin Inhibition versus Enalapril in Heart Failure; PARAMOUNT = Prospective comparison of ARNI with ARB on Management Of heart failUre with preserved ejectioN fracTion; PEERLESS-HF = Prospective evaluation of elastic restraint to lessen the effects of heart failure; QoL = quality of life; RCT = randomized controlled trial; RELAX-AHF = Serelaxin, recombinant human relaxin-2, for treatment of acute heart failure; RESOLVD Pilot Study = Randomized Evaluation of Strategies for Left Ventricular Dysfunction; SOCRATES-REDUCED =Rationale and design of the SOluble guanylate Cyclase stimulatoR in heArT failurE Studies (SOCRATES); SURVIVE = Levosimendan vs dobutamine for patients with acute decompensated heart failure; TOPCAT = Spironolactone for Heart Failure with Preserved Ejection Fraction; Val-HeFT = Valsartan Heart Failure Trial; VNS = vagal nerve stimulation; VO_2 = ventilatory oxygen uptake.

HEART FAILURE WITH PRESERVED EJECTION FRACTION

Heart failure with preserved ejection fraction (HF-PEF) affects many patients with HF (>50% in some series) (22). Despite its high prevalence and growing incidence, evidence-based treatments that have been indisputably shown to reduce morbidity-mortality in this disease are still lacking, and effective therapies are urgently needed (23).

Despite the lack of proven effective therapies, a significant decrease in NT-proBNP in HF-PEF patients has been associated with improved outcomes. In a post hoc analysis from the I-Preserve trial, a 1,000 pg/ml decrease in NT-proBNP from baseline was associated with a reduction in the risk of CV death or HF hospitalization (HR: 0.73; 95% CI: 0.53 to 1.02), and a 1,000 pg/ml increase was associated with an increase in risk (HR: 2.01; 95% CI: 1.50 to 2.69). Beyond a 1,000 pg/ml rise or fall, there was little additional change in risk (24).

In the PARAMOUNT (The angiotensin receptor neprilysin inhibitor LCZ696 in heart failure with preserved ejection fraction: a phase 2 double-blind randomised controlled trial) trial (25), LCZ696 significantly reduced NT-proBNP at 12 weeks compared with valsartan (LCZ696 baseline = 783 pg/ml [IQR: 670 to 914 pg/ml], LCZ696 12 weeks = 605 pg/ml [IQR: 512 to 714 pg/ml]; valsartan baseline = 862 pg/ml [IQR: 733 to 1,012 pg/ml], valsartan 12 weeks = 835 pg/ml [IQR: 710 to 981 pg/ml]; ratio LCZ696/valsartan = 0.77 [95% CI: 0.64 to 0.92], p = 0.005). The PARAMOUNT study was a phase II trial, hence these NT-proBNP differences were not associated with outcomes. Data on CV morbidity and mortality will be provided by the ongoing PARAGON-HF (Efficacy and Safety of LCZ696 Compared to Valsartan, on Morbidity and Mortality in Heart Failure Patients With Preserved Ejection Fraction) trial. Nonetheless, the drop in NPs in HF-PEF is also likely to be associated with lower event rates; however, it is not possible to determine the association between the drop in NPs and a favorable treatment effect due to the lack of indisputably positive trials to date.

ACUTE HEART FAILURE

The association between NPs and outcome in acute heart failure (AHF) has been well documented. Both discharge and change in NP values during hospitalization are strong predictors of subsequent outcomes in AHF patients (10,26,27). However, the association with treatment effect has been more challenging, as demonstrated by the following examples from landmark AHF trials.

A retrospective analysis from the SURVIVE (Levosimendan vs dobutamine for patients with acute decompensated heart failure) trial, a double-blind randomized controlled trial (RCT) comparing levosimendan to dobutamine in patients with severe AHF, found that patients with any BNP reduction (in percent change from baseline to day 5 of hospitalization) had lower all-cause mortality rates as compared with patients who had a BNP increase. Similar findings were observed for patients with an absolute BNP value at day 5 <800 pg/ml. The levosimendan group had greater decreases in BNP level at 24 h (mean change around -631 pg/ml in the levosimendan

group vs. -397 pg/ml in the dobutamine group; $p < 0.001$), which persisted through day 5 ($p < 0.001$ for all time points). However, there were no statistical differences between treatment groups for the primary outcome of all-cause mortality at 180 days (HR: 0.91; 95% CI: 0.74 to 1.13; $p = 0.40$) or any other secondary endpoints of all-cause mortality at 31 days, number of days alive and out of the hospital, patient global assessment, patient assessment of dyspnea at 24 h, and CV mortality at 180 days (28,29) (**Table 1**). The most likely explanation for these findings is the pharmacological difference between the study drugs, as levosimendan has a half-life of approximately 80 h, in contrast to the very short half-life of dobutamine. Hence, at day 5 patients in the levosimendan group had a trend to lower all-cause mortality rates (HR: 0.72; 95% CI: 0.44 to 1.16), but these putative differences had dissipated by days 31 and 180, as shown in the preceding text.

In the RELAX-AHF (Serelaxin, recombinant human relaxin-2, for treatment of acute heart failure) trial, the serelaxin group had a significant decrease in NT-proBNP levels from baseline to day 2 as compared with placebo (geometric mean change = 0.607 in placebo vs. 0.492 in serelaxin, $p < 0.001$; and 30% NT-proBNP decrease from baseline to day 2 = 58.0% in placebo vs. 69.0% in serelaxin, $p < 0.001$). Serelaxin also improved the visual analogic scale (VAS) compared with placebo, but had no significant effect on the other primary endpoint Likert scale. Additionally, no significant effects were recorded for the secondary endpoints of days alive out of the hospital up to day 60, and CV death, readmission for HF, or renal failure until day 180 (HR: 1.02; 95% CI: 0.74 to 1.41; $p = 0.89$). Of note, fewer all-cause deaths at day 180 were observed in the serelaxin group: 65 in the placebo group versus 42 in the serelaxin group (HR: 0.63; 95%CI: 0.42 to 0.93; $p = 0.019$) (30,31). The interpretation of these data is challenging. Serelaxin was administered for 48 h upon admission and, at this time point, it induced a more pronounced BNP drop. However, the coprimary endpoint of the Likert scale and the pre-specified secondary endpoints were not affected, with the exception of 180-day all-cause death (lower in the serelaxin group) (**Table 1**). A confirmatory trial, RELAX-AHF2 (Efficacy, Safety and Tolerability of Serelaxin When Added to Standard Therapy in AHF), designed to assess the effect of serelaxin on morbidity and mortality endpoints, is currently underway (NCT02064868).

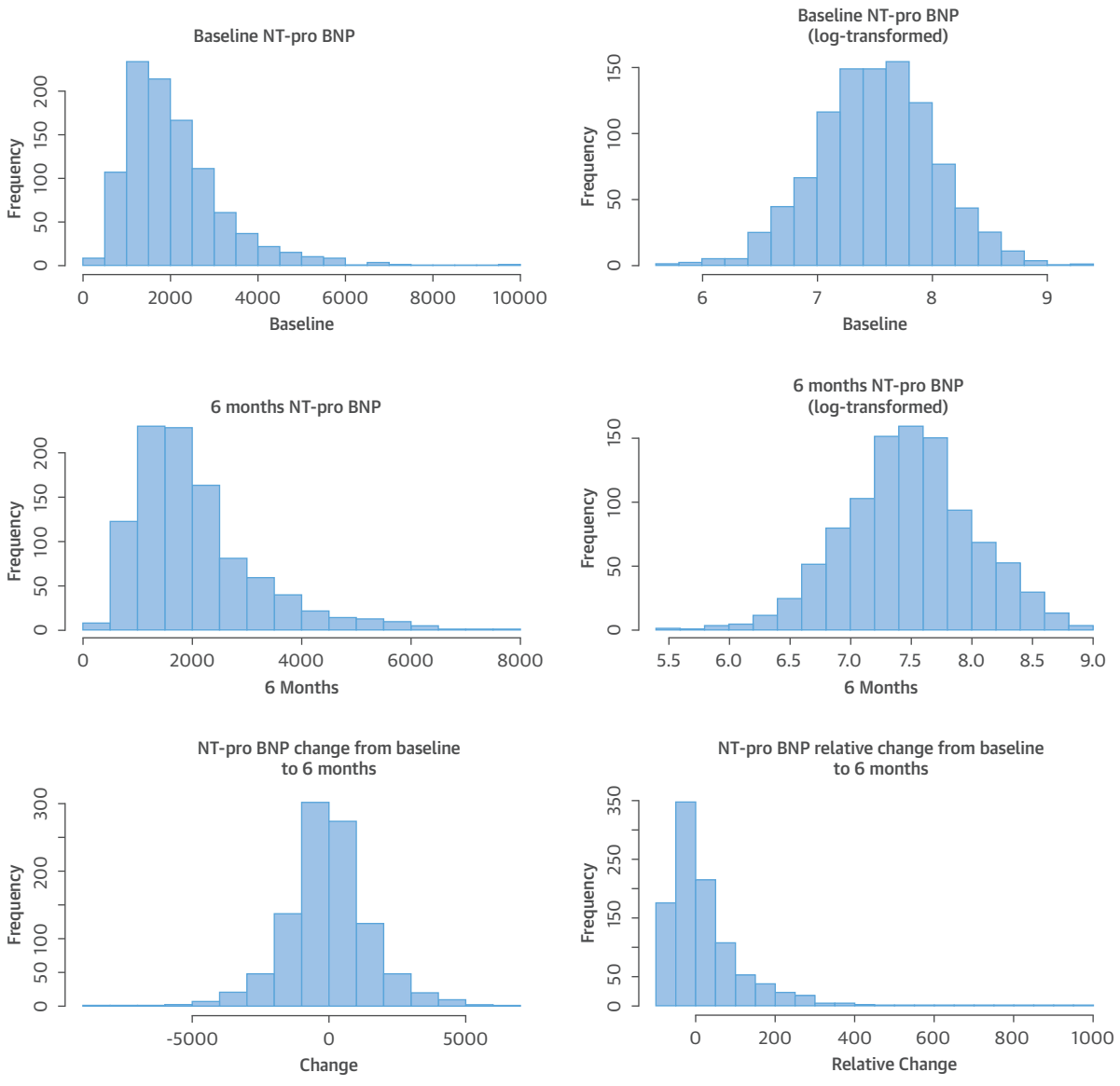
NP CHANGES AND APPLICATION IN TRIALS

Generally, regardless of the method used to assess changes in NP levels in clinical trials, this specific endpoint should be pre-specified in the trial protocol. This allows the mitigation of bias inherent to post hoc analysis and data manipulation.

In general, NPs are skewed variables with high dispersion of values (large IQRs). Hence, the most robust test to compare the differences between groups is a nonparametric test, such as the Mann-Whitney *U* test, or NPs can be log-transformed to "normalize" the distribution, which allows the application of a parametric test, such as a Student *t* test (**Figure 1**).

Given the large dispersion of NP values, calculation of NP changes adjusted on baseline values can also provide useful information, once pre-specified, even though, in a randomized trial, this "baseline adjustment" is theoretically secured by the randomization process itself, potentially eliminating allocation bias. Hence, the median difference in NP or mean difference in log-NP are likely to be the most robust methods to detect "clinically significant" differences between groups. Percent changes can also be used, as NP reductions relative to baseline are associated with improved outcomes (as in the COPERNICUS trial, described previously). The most relevant information is that the formula used to detect NP differences is pre-specified. This method can avoid bias induced by different formulas and computation methods. In **Table 2**, we provide a hypothetical example where 1,000 patients were enrolled in a putative trial: 500 assigned to placebo and 500 assigned to an active treatment. This hypothetical example was built using simulated data with a log-normal distribution for both baseline treatment and placebo, and also for placebo at 6 months. For the treatment group at 6 months, we modified the parameters of the log-normal distribution to obtain a small reduction in a random fashion. At baseline, there were no differences between groups in the NT-proBNP values (NT-proBNP = 2,040.4 ± 1,048.5 pg/ml in the placebo group vs. 2,097.4 ± 1,133.3 pg/ml in the treatment group; $p = 0.409$), whereas at 6 months, significant differences were detected with a nonparametric test or log transformation, but not with a parametric test (NT-proBNP = 1,858.3 pg/ml [IQR: 1,329.8 to 2,509.8 pg/ml] in the placebo group vs. 1,737.1 pg/ml [IQR: 1,206.5 to 2,451.1 pg/ml] in the treatment group, $p = 0.045$ with a nonparametric test; and NT-proBNP = 2,092.3 ± 1,098.6 pg/ml in the placebo group vs. 2,001.6 ± 1,141.1 pg/ml in treatment group, $p = 0.200$ with a parametric test) (**Table 2**). Of note, NT-proBNP had a right-skewed distribution; hence, a nonparametric test or log transformation is more appropriate (**Figure 1**). The ratio of geometric means also provided different results from those obtained with arithmetic means, with the former providing a

FIGURE 1 **NT-proBNP Distribution According to Absolute, Log-Transformed, Delta, and Relative Delta Values**



The use of crude, log-transformed, absolute change, or relative change values alters the distribution of the variable, potentially influencing result interpretation. NT-proBNP = N-terminal pro-B-type natriuretic peptide.

p value closer to statistical significance ($p = 0.0511$ with ratio of geometric means vs. $p = 0.134$ with arithmetic means) (Table 2). The proportion of patients achieving a 30% reduction in NT-proBNP values also provided a totally different result with no significant differences between groups (28.4% in the placebo group vs. 31.2% in the treatment group; $p = 0.369$).

In conclusion, either absolute or percent changes are likely to be appropriate for determining clinically significant differences between treatment groups in HF-REF. However, these methods need to be prespecified to avoid post hoc adaptations.

DROP IN NP LEVELS: HOW MUCH IS CLINICALLY MEANINGFUL?

The examples shown in the previous section demonstrate that the methods and partition values used to determine differences between NP levels vary

TABLE 2 Comparison of Different Formulas Frequently Used to Calculate Changes in Natriuretic Peptides (NT-proBNP)

Time Points	Test/Values	Placebo (n = 500)	Treatment (n = 500)	p Value
Baseline	Parametric test	2040.4 ± 1048.5	2097.4 ± 1133.3	0.409
	Nonparametric test	1778 (1301.4.2 to 2575.7)	1906.8 (1304.8 to 2553.8)	0.456
	Geometric mean	1803.185	1844.733	–
Log (baseline)	Parametric test	7.497 ± 0.503	7.520 ± 0.509	0.477
	Nonparametric test	7.5 (7.2 to 7.9)	7.6 (7.2 to 7.8)	0.456
6 months	Parametric test	2092.3 ± 1098.6	2001.6 ± 1141.1	0.200
	Nonparametric test	1858.3 (1329.8 to 2509.8)	1737.1 (1206.5 to 2451.1)	0.045
	Geometric mean	1848.474	1730.298	–
Log (6 months)	Parametric test	7.522 ± 0.499	7.456 ± 0.543	0.045
	Nonparametric test	7.5 (7.1 to 7.8)	7.5 (7.1 to 7.8)	0.045
Change from baseline (ratio of geometric means: 6 months/baseline)	–	1.025	0.938	–
Difference of means (treatment–placebo, log-transformed)	-0.089	–	–	0.0511
Ratio of geometric means (treatment group change from baseline/placebo group change from baseline)	0.915	–	–	0.0511
Absolute delta: 6 months–baseline	Parametric test	51.9 ± 1539.7	-95.9 ± 1575.5	0.134
	Nonparametric test	-29.3 (-806.9 to 891.7)	-112.9 (-885.3 to 690.3)	0.208
Relative delta: 6 months–baseline, %	Parametric test	35 ± 122	21 ± 101	0.0508
	Nonparametric test	-2 (-34 to 63)	-8 (-42 to 49)	0.086
Proportion of patients with >30% reduction, %	Chi-square test	28.4	31.2	0.369

Values are mean ± SD or median (interquartile range).
NT-proBNP = N-terminal pro-B-type natriuretic peptide.

widely. In addition to the variation in the methods used to compute the change in NPs, assay variability (analytical and biological) should also be acknowledged. In "healthy subjects," NPs are usually within the lower threshold for detection; hence, small absolute changes without biological significance may result in an apparently high percent variation (32). However, in symptomatic HF patients, NP levels may present larger absolute changes (even if the percent variation is low), leading some authors to suggest that only variations 3 times greater than the SD of the assay variability should be regarded as clinically significant (33). In severely symptomatic HF patients, NP levels are usually somewhat above the reference range, and, in the setting of a double-blind RCT, these assays are subject of random analytical and biological variability (i.e., nonsystematic error). In other words, in the case of a double-blind RCT, NP variability will also be subject to the "blinded random effect," varying within and between patients and assays, and hence increasing the strength of the association between the exposure and the outcome (34). If prespecified statistical differences between blinded treatment groups are found, they are also likely to be clinically relevant.

For the sake of reproducibility, in "positive" HF-REF trials (e.g., PARADIGM-HF and CARE-HF trials), the NT-proBNP levels were lower in the treatment group using mean or median value comparisons. In others, the percent drop has also been associated with improved outcomes (e.g., the AREA IN-CHF and COPERNICUS trials). These associations seem to be independent of "partition values," and any statistically significant difference using these simple statistical methods is likely to indicate a clinically beneficial treatment effect over a comparator. For the sake of power calculation, the use of a predetermined target NP drop may be needed. Therefore, a ≥30% reduction from baseline is a reasonable compromise in HF-REF trials. This magnitude of change in NP levels is usually associated with a robust benefit on clinical outcomes (**Table 1**).

6-MIN HALL WALK DISTANCE TEST. The 6MWD is an inexpensive and reproducible method to assess exercise tolerance. It can be performed in a majority of HF patients, even when exercise capacity is limited by severity of disease or multiple comorbidities. Results of the 6MWD have been correlated with exercise capacity measured by formal treadmill and cardiopulmonary exercise tests, and have also been associated with morbidity and mortality in HF patients (35,36).

HF-REF. The 6MWD has been used in a large number of HF trials. The 6MWD protocol is standardized, and associations between changes in 6MWD distance and outcomes or treatment effects have been provided. In and of itself, the 6MWD represents a clinically

relevant measure of patient symptom status. When performed using a standard protocol, including study personnel blinding and at least 2 pre-randomization studies, 6MWD represents a reasonably objective method to quantitate exercise capacity and reflects a clinical issue important to patients with HF. The objectivity and reproducibility of the 6MWD is enhanced when patients and investigators are blinded, and when device trials implant both the control and treatment arms. However, these designs are not always practical.

A systematic review of the 6MWD as an outcome measure for the assessment of treatment effects in chronic HF RCTs has been performed. The 6MWD improved in the majority of trials of CRT, but showed variability in some pharmacological (e.g., angiotensin-converting enzyme inhibitors and beta-blockers) and device (e.g., VNS) trials. Of note, 6MWD improved with ibopamine treatment, an intervention that was demonstrated to be deleterious (37).

A more recent meta-analysis showed that only ≈50% of the trials that tested therapies with positive effects on morbidity and mortality showed a corresponding significant increase in the 6MWD distance (4). An agreement between increased 6MWD and survival benefit was more consistently observed in large CRT trials including severe and symptomatic HF-REF patients (38–41).

Further examples from large RCTs support the association between 6MWD improvement and a favorable treatment effect. In a CRT trial by Cazeau et al. (42), the mean distance in the 6MWD was ~22% greater after 3 months with active pacing (399 ± 100 m vs. 326 ± 134 m; $p < 0.01$). In this trial, QoL and peak oxygen consumption also improved and, more importantly, HF hospitalizations were reduced. In another trial by Abraham et al. (43), patients assigned to CRT experienced an improvement of ~26% in the 6MWD (+39 m vs. +10 m; $p < 0.01$). In this trial, QoL, time on the treadmill, and LVEF also improved. However, in the MIRACLE-ICD (Combined cardiac resynchronization and implantable cardioversion defibrillation in advanced chronic heart failure) I and II trials, patients assigned to CRT did not improve in 6MWD as compared with controls (55 m vs. 53 m; $p = 0.36$). However, CRT improved QoL, functional status, and exercise capacity in patients with moderate to severe HF, a wide QRS interval, and life-threatening arrhythmias (41). The INOVATE-HF (Vagus Nerve Stimulation for the Treatment of Heart Failure) trial (44) tested the efficacy and safety of VNS in symptomatic HF-REF. The primary efficacy endpoint was the composite of death from any cause or first event for worsening HF. In the INOVATE-HF trial, VNS did not improve the primary efficacy outcome (HR: 1.14; 95% CI: 0.86 to 1.53; $p = 0.37$). However, VNS did improve 6MWD distance (+28.2 m in the VNS group vs. -4.6 m in the control group; $p < 0.01$). Notably, no statistically significant differences between groups were found regarding left ventricular end-systolic volume index (-5.4 ml/m^2 in the VNS group vs. -2.8 ml/m^2 in the control group; $p = 0.49$) (**Table 1**). However, in other neuromodulatory studies, such as the HOPE4HF study, there were correlations between improvement in 6MWD and HF hospitalization rate. BAT increased 6MHW distance by 58.1 ± 19.8 m ($p = 0.004$) compared with standard medical treatment. This was associated with a 52% relative reduction in the number of HF hospitalizations per year and an 82% reduction in HF hospitalizations days per year (21).

The variability of 6MWD distance, and its correlation with morbidity and mortality are not surprising because the 6MWD is itself a reflection of exercise tolerance, limited by several non-CV factors, such as conditioning, osteoarticular pathology, patient effort, and willingness/motivation to perform the test. In addition, the 6MWD (and other exercise parameters) also rely on the ability of skeletal muscle to extract oxygen from blood, pulmonary and endothelial function, and cardiac output (45). Moreover, the 6MWD is likely to perform better (as prognostic tool) in patients with severe and symptomatic HF whose 6MWD is most severely limited, and an improvement could be clinically meaningful (46).

However, in the absence of any other effective therapy, the increase in 6MWD distance, in agreement with NP, drops. In this case, cardiac structural improvement may be considered a clinically meaningful favorable treatment response and, when associated with reasonable safety data, may support its approval or testing in a phase III trial.

INCREASE IN 6MWD: HOW MUCH IS ENOUGH?

There are limited data to define how much of a difference between treatment groups in 6MWD distance is enough to define a test as “positive” or “indicative of efficacy.” However, an increase in 6MWD is of much more value to a severely symptomatic patients. For example, a 50-m increase in 6MWD is of greater value to a patient with a baseline 6MWD of 100 m compared with a patient with a baseline of 300 m. In general, a 30- to 50-m increase in 6MWD is considered a clinically significant improvement, is associated with significant improvement in NYHA functional class and health-related QoL, and has been used in the CRT trials as relevant to pre-market approval.

From the described trials, we may observe that in device trials, 6MWD, treatment effect, and outcomes more consistently agreed, when patients had severe disease and symptoms. The 6MWD should be used in similar settings, and a statistically significant difference between mean distances or absolute delta distances should be considered as "pointing toward efficacy."

Importantly, 6MWD should be used in collaboration with other intermediate endpoints. As stated earlier, a positive effect on 6MWD is more valuable when it is congruent with positive changes in NP levels, or in another remodeling assessment, such as echocardiography or magnetic resonance imaging parameters (please see also the Practical Suggestions for Trials section).

QoL. Health-related QoL questionnaires have been used extensively in cardiology research; however, data regarding external validation and association with prognosis are somewhat limited. Nonetheless, poor scores on QoL questionnaires have been associated with a dismal prognosis in HF (47,48).

HF-REF. QoL has also been used as a clinically relevant endpoint and an intermediate prognostic marker in several HF-REF trials. The most commonly used quality instruments are the Minnesota Living with Heart Failure Questionnaire (MLHFQ) and the Kansas City Cardiomyopathy Questionnaire (KCCQ). In HF-REF with severe symptoms, an improvement in health-related QoL also points toward a beneficial and clinically meaningful treatment effect if aligned with the other tests, particularly NPs and cardiac structural measures. Some examples are provided in the following text.

In the Cazeau et al. (42) and Abraham et al. (43) trials described earlier, active pacing improved QoL, as assessed by the MLHFQ (-18 vs. -9 points, $p < 0.01$, in Abraham et al. [43]). In another trial by Auricchio et al. (49), CRT in patients with a QRS interval >150 ms improved QoL, as assessed by the MLHFQ (17 ± 13 points vs. 25 ± 15 points; $p < 0.01$), and CRT significantly improved the maximal and submaximal exercise capacity, 6MWD, and functional status of HF patients, especially those with QRS intervals >150 ms (49). The same was observed in Higgins et al. (39), in which CRT improved QoL from baseline to 6 months compared with the no CRT group, as assessed by the MLHFQ (-17.5 vs. -11.0, $p = 0.02$). In this trial, CRT also improved functional status in patients indicated for an implantable cardioverter-defibrillator (ICD) who also had symptomatic HF and an intraventricular conduction delay (39). In each of these trials, treatment was associated with improved morbidity and mortality, suggesting an association between change in QoL and those outcomes.

In a trial by Toblli et al. (50), iron treatment improved QoL, as assessed by the MLHFQ (59 ± 8 points vs. 41 ± 7 points; $p < 0.05$) compared with placebo at 6 months. Iron therapy also improved functional capacity, symptoms, and NT-proBNP (50). In the BEST (Beta-Blocker Bucindolol in Patients with Advanced Chronic Heart Failure) trial evaluating bucindolol versus placebo, QoL questionnaires were strongly associated with all-cause mortality in the overall study population. However, at 12 months, bucindolol-treated patients did not have a QoL improvement according to the MLHFQ (51) or better outcomes. In the MIRACLE ICD II trial, CRT also did not improve QoL scores as compared with ICD in NYHA functional class II patients and those with QRS intervals >130 ms. However, CRT improved cardiac structure and function, and the composite clinical response over 6 months (**Table 1**) (38). In an analysis of the HF-ACTION (Efficacy and Safety of Exercise Training in Patients With Chronic Heart Failure) trial, Flynn et al. (52) reported a modest, but sustained improvement in the KCCQ score in the exercise arm of the trial. In the main trial, in the protocol-specified primary analysis, exercise training consistently reduced the primary endpoint of all-cause mortality or hospitalizations for any cause in baseline-adjusted analyses (53). In 2 studies of iron replacement in patients with HF-REF, the KCCQ score improvement agreed with the primary outcome. Anker et al. (54) reported that intravenous iron replacement improved the primary endpoint of patient global assessment and NYHA functional class at 24 weeks in the FAIR-HF (Ferric Carboxymaltose in Patients with Heart Failure and Iron Deficiency) trial. There were clear improvements in KCCQ score at 4, 12, and 24 weeks in this study, concordant with the functional improvement. In the CONFIRM-HF (Beneficial effects of long-term intravenous iron therapy with ferric carboxymaltose in patients with symptomatic heart failure and iron deficiency) trial, Ponikowski et al. (55) also demonstrated a significant improvement in the primary endpoint of increased 6MWD at 24 weeks, which agreed with a significant improvement in the KCCQ score. In a report of the first 122 patients enrolled in the PEERLESS (HeartNet cardiac restraint device, Paracor Medical Inc., Sunnyvale, California) restraint device, Constanzo et al. (56) reported an improvement in KCCQ score in the device versus the control group. Although, the study was stopped due to the prediction of futility in the primary endpoint of ventilatory oxygen uptake (VO_2) in the device group, the secondary ventricular remodeling endpoints were positive and

correlated with the improvements in the KCCQ. In the MADIT-CRT (Cardiac-Resynchronization Therapy for the Prevention of Heart-Failure Events) trial, the device group had lower rates of the primary endpoint of all-cause death and HF events (57). Veazie et al. (58) reported that the KCCQ improved only in subjects with an underlying left bundle branch block, but the largest outcome benefit in MADIT-CRT was also observed in patients with a left bundle branch block. In the INOVATE-HF (Vagus Nerve Stimulation for the Treatment of Heart Failure) trial, VNS also improved QoL as assessed by the KCCQ score (+11.2 points in the VNS group vs. +6.9 points in the control group; $p < 0.01$). However, as described previously, in this trial, VNS therapy did not improve the primary outcome of all-cause death or HF hospitalization (**Table 1**). However, as described previously for 6MWD, in the HOPE4HF study, an agreement between improvement in MLHFQ QoL (BAT decreased MLHFQ scores -19.5 ± 4.2 points vs. standard medical treatment; $p < 0.001$) and rate of HF hospitalization was observed.

Similar to the 6MWD, the agreement between the effects of treatment on health-related QoL and outcomes has some variability. The subjectivity of QoL questionnaires may play a role; patients' motivation, other comorbidities (e.g., depression, dementia), and social support or financial conditions may play an important role in QoL questionnaire results. In concordance with the 6MWD, QoL questionnaires are likely to perform better in severe and symptomatic HF-REF patients, and point toward a beneficial effect, especially when the QoL improvement is consistent with other tests (NPs, imaging, and 6MWD) (**Table 1**).

HF-PEF. As described previously, the absence of an unequivocally positive trial does not allow us to evaluate the role of health-related QoL questionnaires as treatment effect surrogates in HF-PEF. The TOPCAT (Spironolactone for Heart Failure with Preserved Ejection Fraction) trial was a large, international RCT evaluating the use of spironolactone versus placebo in HF-PEF (59). Spironolactone did not significantly reduce the primary composite outcome of death from CV causes or hospitalizations for HF (HR: 0.89; 95% CI: 0.77 to 1.04; $p = 0.14$). However, Lewis et al. (60) reported a close correlation of improvement in KCCQ and the use of spironolactone. The results of the TOPCAT trial are challenging, as post hoc analysis has revealed subgroups that are likely to experience a beneficial effect with spironolactone (61–64). Therefore, the interpretation of the reliability of a change in the QoL scores as surrogate markers in HF-PEF still needs further assessment.

INCREASE IN QoL PARAMETERS: HOW MUCH IS ENOUGH?

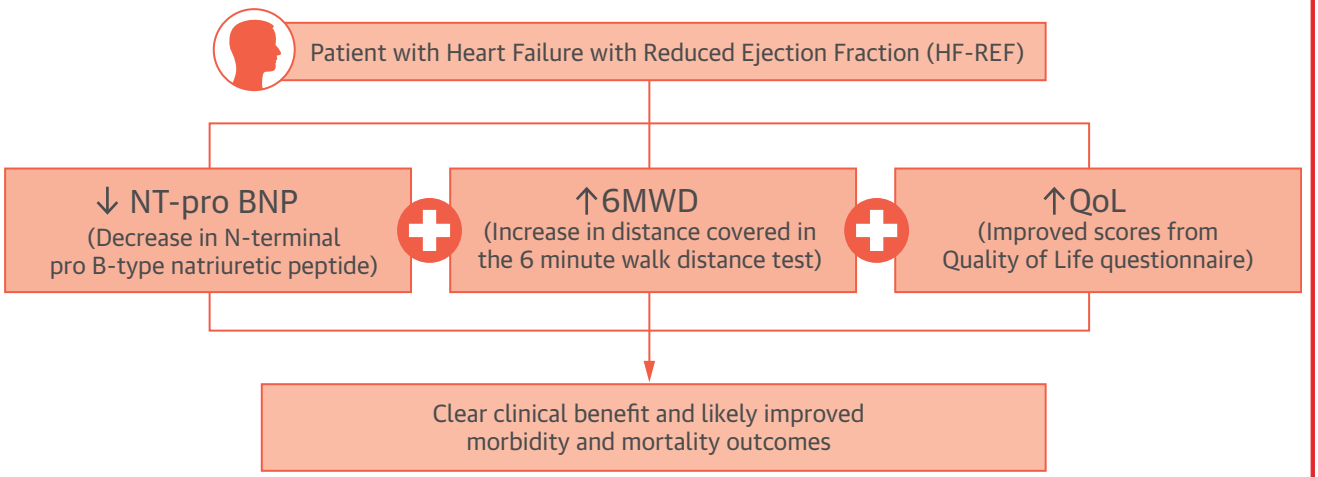
There is little data to define how much QoL improvement is associated with treatment efficacy. In severe and symptomatic HF-REF, a statistically significant difference between treatment allocation groups can be regarded as potentially beneficial. In general, a 5- to 10-point decrease in MLHFQ QoL score is considered clinically and prognostically significant, and more so, if, as mentioned previously, this finding is congruent with decreased NP level and improved 6MWD.

PRACTICAL SUGGESTIONS FOR TRIALS

HF-REF. The use of NPs should be encouraged as the primary efficacy outcome in severe and symptomatic HF-REF trials. The NP change between treatment groups should be regarded as the most robust predictor of efficacy, especially if accompanied by improvement in cardiac structure and function measurements. In this regard, the echocardiographic assessment of left ventricular end-systolic volume (as a reproducible and easily available left ventricular reverse-remodeling parameter) is inversely associated with outcome and response to device and pharmacological therapy (65–67). The structural pattern of left ventricular remodeling and evidence of scarring on cardiac magnetic resonance also has demonstrated prognostic value (68). However, it is a resource-demanding method unlikely to be broadly performed in a multicenter RCT.

With the advent of angiotensin-neprilysin inhibition for the treatment of symptomatic HF-REF (69), the use of NPs to monitor treatment response has been debated (70). The enzyme neprilysin efficiently degrades circulating BNP in vivo, whereas proBNP and NT-proBNP are virtually resistant to enzymatic cleavage. It seems plausible that measuring BNP in patients taking neprilysin inhibitors, such as LCZ696, may not reliably reflect cardiac dysfunction (70). Moreover, BNP variability under neprilysin inhibition also depends on the assay used (71). In summary, BNP levels are likely to increase under neprilysin inhibition, whereas NT-proBNP levels decrease in a sustained fashion. In other words, BNP will reflect the action of the drug, whereas levels of NT-proBNP will reflect the effects of the drug on the heart (19). Hence, BNP can potentially be used for compliance monitoring, and NT-proBNP for treatment effect assessment. Consequently, in future trials, NT-proBNP will likely become the preferred NP for monitoring treatment response (e.g., device trials).

CENTRAL ILLUSTRATION **Clinically Meaningful Endpoints in Heart Failure Trials: HF-REF, HF-PEF, and AHF**



Ferreira, J.P. et al. J Am Coll Cardiol. 2016;68(24):2690–707.

In the case of heart failure with preserved ejection fraction (HF-PEF) and acute heart failure (AHF), the agreement between the clinically meaningful endpoints and the treatment effect is not consistent, and is discussed in detail in the corresponding sections of the paper. 6MWD = 6-min walking test distance; HF-REF = heart failure with reduced ejection fraction; M&M = morbidity and mortality; NT-proBNP = N terminal pro-B-type natriuretic peptide; QoL = health-related quality of life.

The 6MWD and QoL questionnaires are also associated with treatment effects. They should be used together with other intermediate endpoints in HF trials. A positive effect on 6MWD or QoL has significant value when it is congruent with positive changes in NP levels, or another remodeling assessment. More generally, consistent changes in NPs, 6MWD, and QoL in a phase II trial may be taken as strongly indicative of potential clinical benefit (Central Illustration), and are encouraging trends for drug/device approval (in the absence of effective alternatives) or undertaking an outcome trial. In the context of the EAP program, significant and concordant improvement in these 3 intermediate endpoints, together with clear indications of safety and directional signals toward reductions in morbidity and mortality, constitute premarket approval application-level data.

HF-PEF AND AHF. The absence of an unequivocally positive trial presents a challenge to evaluating the role of intermediate endpoints in determining efficacy of new treatment and management approaches in patients with HF-PEF. However, the characteristics of this patient population, the rapidly increasing prevalence rate, the extraordinarily high mortality rate, and the profound disability experienced by HF-PEF patients, have created an absolute mandate that novel approaches to RCT design in both phases II and III must be developed. The use of intermediate endpoints, such as NPs, 6MWD, QoL, and other novel approaches, represent an important unmet need in CV medicine. These considerations are also applicable for management of AHF.

FUTURE DIRECTIONS. As we may observe from the many examples provided, there is an urgent need for more robust markers of efficacy. These markers should be reproducible and resist bias. In addition, novel approaches should be examined in combination with other endpoints to obtain congruency, assessing several physiopathological pathways, and maximizing internal and external validity.

In addition to NPs, future trials assessing the effect of a given therapy on biomarkers of cardiac injury (e.g., high-sensitivity troponin T), inflammation or remodeling or fibrosis (e.g., collagen metabolism markers, Galectin-3 and soluble ST2), and cardiac function (e.g., corin) and also on "novel" imaging left ventricular reverse-remodeling parameters (e.g., assessed by echocardiographic left ventricular strain or magnetic resonance), may help build the mechanistic plausibility of other clinically meaningful intermediate endpoints.

Precompetitive and academic consortiums should further examine whether intermediate endpoint data can be used to improve the value and trial applicability of intermediate endpoints. Moreover, adaptive trial design has been proposed as a means to increase

the efficiency of randomized clinical trials, potentially benefiting trial participants and future patients, while reducing costs and enhancing the likelihood of finding a true benefit, if one exists, of the therapy being studied (72). In this regard, Bayesian adaptive trial designs are particularly well suited to studies with a primary clinical endpoint and related intermediate endpoints. During trial updates, interim data are analyzed using a statistical model that estimates the treatment effect(s), as well as the strength of the relationship between the primary and intermediate endpoints. Provided that early trial data confirm that the intermediate endpoints are closely related to primary outcomes, the intermediate endpoints can assist in predicting the outcome of the eventual primary analysis, potentially reducing the number of patients enrolled and improving the study conclusions (73,74).

The development of trials with a single master protocol in which multiple treatments are evaluated simultaneously ("platform trials"), could offer flexible features, such as dropping treatments for futility, declaring 1 or more treatments superior, or adding new treatments to be tested during the course of a trial, requiring fewer patients and less time (74,75). Data envelopment analysis and Bayesian adaptive designs have potential to be used to estimate the relative performance of competing designs in the presence of multiple criteria (76).

CONCLUSIONS

In symptomatic HF-REF, statistically significant absolute and relative decreases in NPs can be taken as objective mechanistic and prognostic markers indicative of clinical benefit and are frequently associated with improved outcomes. When congruent with statistically significant improvements in patient-centered outcomes, 6MWD, and health-related QoL scores, these changes are reasonable evidence of clinically meaningful benefit of a given treatment. We suggest considering such endpoints in combination when assessing medical device therapy in the novel FDA EAP. Our review of published reports provides evidence to support the use of 3 such intermediate endpoints in severe HF-REF.

REPRINT REQUESTS AND CORRESPONDENCE: Dr. Faiez Zannad, Centre d'Investigation Clinique Inserm, CHU, Université de Lorraine, Institut Lorrain du Coeur et des Vaisseaux, 4, rue du Morvan, 54500 Vandoeuvre-Les-Nancy, France. E-mail: f.zannad@chru-nancy.fr.

REFERENCES

1. Greene SJ, Gheorghiade M. Same protocol, different continents, different patients: should we continue to conduct global heart failure trials? Eur J Heart Fail 2015;17:875–8.

2. Food and Drug Administration. FDA-2014-D-0363: Expedited Access for Premarket Approval and De Novo Medical Devices Intended for Unmet Medical Need for Life Threatening or Irreversibly Debilitating Diseases or Conditions. Rockville, MD: Food and Drug Administration, 2015. Available at: http://www.fda.gov/downloads/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/UCM393978.pdf. Accessed September 29, 2016.

3. Masson S, Latini R, Anand IS, et al., Val-HeFT Investigators. Prognostic value of changes in N-terminal pro-brain natriuretic peptide in Val-HeFT (Valsartan Heart Failure Trial). J Am Coll Cardiol 2008;52:997–1003.

4. Wessler BS, Kramer DG, Kelly JL, et al. Drug and device effects on peak oxygen consumption, 6-minute walk distance, and natriuretic peptides as predictors of therapeutic effects on mortality in patients with heart failure and reduced ejection fraction. Circ Heart Fail 2011;4:578–88.

5. Seino Y, Ogawa A, Yamashita T, et al. Application of NT-proBNP and BNP measurements in cardiac care: a more discerning marker for the detection and evaluation of heart failure. Eur J Heart Fail 2004;6:295–300.

6. Groenning BA, Raymond I, Hildebrandt PR, et al. Diagnostic and prognostic evaluation of left ventricular systolic heart failure by plasma N-terminal pro-brain natriuretic peptide concentrations in a large sample of the general population. Heart 2004;90:297–303.

7. Maeda K, Tsutamoto T, Wada A, et al. Plasma brain natriuretic peptide as a biochemical marker of high left ventricular end-diastolic pressure in patients with symptomatic left ventricular dysfunction. Am Heart J 1998;135:825–32.

8. Kazanegra R, Cheng V, Garcia A, et al. A rapid test for B-type natriuretic peptide correlates with falling wedge pressures in patients treated for decompensated heart failure: a pilot study. J Card Fail 2001;7:21–9.

9. Kallistratos MS, Dritsas A, Laoutaris ID, et al. N-terminal prohormone brain natriuretic peptide as a marker for detecting low functional class patients and candidates for cardiac transplantation: linear correlation with exercise tolerance. J Heart Lung Transplant 2007;26:516–21.

10. Bettencourt P, Azevedo A, Pimenta J, et al. N-terminal-pro-brain natriuretic peptide predicts outcome after hospital discharge in heart failure patients. Circulation 2004;110:2168–74.

11. Anand IS, Fisher LD, Chiang YT, et al. Changes in brain natriuretic peptide and norepinephrine over time and mortality and morbidity in the Valsartan Heart Failure Trial (Val-HeFT). Circulation 2003;107:1278–83.

12. Latini R, Masson S, Wong M, et al. Incremental prognostic value of changes in B-type natriuretic peptide in heart failure. Am J Med 2006;119:70.e23–30.

13. Hartmann F, Packer M, Coats AJ, et al. Prognostic impact of plasma N-terminal pro-brain natriuretic peptide in severe chronic congestive heart failure: a substudy of the Carvedilol Prospective Randomized Cumulative Survival (COPERNICUS) trial. Circulation 2004;110:1780–6.

14. Hartmann F, Packer M, Coats AJ, et al. NT-proBNP in severe chronic heart failure: rationale, design and preliminary results of the COPERNICUS NT-proBNP substudy. Eur J Heart Fail 2004;6:343–50.

15. Savarese G, Musella F, D'Amore C, et al. Changes of natriuretic peptides predict hospital admissions in patients with chronic heart failure: a meta-analysis. J Am Coll Cardiol HF 2014;2:148–58.

16. Cleland JGF, Daubert JC, Erdmann E, et al., Cardiac Resynchronization-Heart Failure (CARE-HF) Study Investigators. The effect of cardiac resynchronization on morbidity and mortality in heart failure. N Engl J Med 2005;352:1539–49.

17. Fruhwald FM, Fahrleitner-Pammer A, Berger R, et al. Early and sustained effects of cardiac resynchronization therapy on N-terminal pro-B-type natriuretic peptide in patients with moderate to severe heart failure and cardiac dyssynchrony. Eur Heart J 2007;28:1592–7.

18. Boccanelli A, Mureddu GF, Cacciatore G, et al., AREA IN-CHF Investigators. Anti-remodelling effect of canrenone in patients with mild chronic heart failure (AREA IN-CHF study): final results. Eur J Heart Fail 2009;11:68–76.

19. Packer M, McMurray JJV, Desai AS, et al., PARADIGM-HF Investigators and Coordinators. Angiotensin receptor neprilysin inhibition compared with enalapril on the risk of clinical progression in surviving patients with heart failure. Circulation 2015;131:54–61.

20. Gheorghiade M, Böhm M, Greene SJ, et al., ASTRONAUT Investigators and Coordinators. Effect of aliskiren on postdischarge mortality and heart failure readmissions among patients hospitalized for heart failure: the ASTRONAUT randomized trial. JAMA 2013;309:1125–35.

21. Abraham WT, Zile MR, Weaver FA, et al. Baroreflex activation therapy for the treatment of heart failure with a reduced ejection fraction. J Am Coll Cardiol HF 2015;3:487–96.

22. Cleland JG, Pellicori P. Defining diastolic heart failure and identifying effective therapies. JAMA 2013;309:825–6.

23. Ferreira JP, Girerd N, Rossignol P, et al. Geographic differences in heart failure trials. Eur J Heart Fail 2015;17:893–905.

24. Jhund PS, Anand IS, Komajda M, et al. Changes in N-terminal pro-B-type natriuretic peptide levels and outcomes in heart failure with preserved ejection fraction: an analysis of the I-Preserve study. Eur J Heart Fail 2015;17:809–17.

25. Solomon SD, Zile M, Pieske B, et al., Prospective comparison of ARNI with ARB on Management Of heart failUre with preserved ejectioN fracTion (PARAMOUNT) Investigators. The angiotensin receptor neprilysin inhibitor LCZ696 in heart failure with preserved ejection fraction: a phase 2 double-blind randomised controlled trial. Lancet 2012;380:1387–95.

26. Logeart D, Thabut G, Jourdain P, et al. Predischarge B-type natriuretic peptide assay for identifying patients at high risk of re-admission after decompensated heart failure. J Am Coll Cardiol 2004;43:635–41.

27. Stienen S, Salah K, Dickhoff C, et al. N-terminal pro-B-type natriuretic peptide (NT-proBNP) measurements until a 30% reduction is attained during acute decompensated heart failure admissions and comparison with discharge NT-proBNP levels: implications for in-hospital guidance of treatment. J Card Fail 2015;21:930–4.

28. Mebazaa A, Nieminen MS, Packer M, et al., SURVIVE Investigators. Levosimendan vs dobutamine for patients with acute decompensated heart failure: the SURVIVE Randomized Trial. JAMA 2007;297:1883–91.

29. Cohen-Solal A, Logeart D, Huang B, et al. Lowered B-type natriuretic peptide in response to levosimendan or dobutamine treatment is associated with improved survival in patients with severe acutely decompensated heart failure. J Am Coll Cardiol 2009;53:2343–8.

30. Metra M, Cotter G, Davison BA, et al., RELAX-AHF Investigators. Effect of serelaxin on cardiac, renal, and hepatic biomarkers in the Relaxin in Acute Heart Failure (RELAX-AHF) development program: correlation with outcomes. J Am Coll Cardiol 2013;61:196–206.

31. Teerlink JR, Cotter G, Davison BA, et al., RELAXin in Acute Heart Failure (RELAX-AHF) Investigators. Serelaxin, recombinant human relaxin-2, for treatment of acute heart failure (RELAX-AHF): a randomised, placebo-controlled trial. Lancet 2013;381:29–39.

32. Richards AM. Variability of NT-proBNP levels in heart failure: implications for clinical application. Heart 2007;93:899–900.

33. Clerico A, Zucchelli GC, Pilo A, et al. Clinical relevance of biological variation of B-type natriuretic peptide. Clin Chem 2005;51:925–6.

34. Zhang F, Knoll A. Systematic error modeling and bias estimation. Sensors (Basel) 2016;16:729.

35. Bittner V, Weiner DH, Yusuf S, et al. Prediction of mortality and morbidity with a 6-minute walk test in patients with left ventricular dysfunction. JAMA 1993;270:1702–7.

36. Guyatt GH, Sullivan MJ, Thompson PJ, et al. The 6-minute walk: a new measure of exercise capacity in patients with chronic heart failure. Can Med Assoc J 1985;132:919–23.

37. Olsson LG, Swedberg K, Clark AL, et al. Six minute corridor walk test as an outcome measure for the assessment of treatment in randomized, blinded intervention trials of chronic heart failure: a systematic review. Eur Heart J 2005;26:778–93.

38. Abraham WT, Young JB, Leon AR, et al. Effects of cardiac resynchronization on disease progression in patients with left ventricular systolic dysfunction, an indication for an implantable cardioverter-defibrillator, and mildly symptomatic chronic heart failure. Circulation 2004;110:2864–8.

39. Higgins SL, Hummel JD, Niazi IK, et al. Cardiac resynchronization therapy for the treatment of heart failure in patients with intraventricular conduction delay and malignant ventricular tachyarrhythmias. J Am Coll Cardiol 2003;42:1454–9.

40. Bristow MR, Saxon LA, Boehmer J, et al., Comparison of Medical Therapy, Pacing, and Defibrillation in Heart Failure (COMPANION) Investigators. Cardiac-resynchronization therapy with or without an implantable defibrillator in advanced chronic heart failure. N Engl J Med 2004;350:2140–50.

41. Young JB, Abraham WT, Smith AL, et al., Multicenter InSync ICD Randomized Clinical Evaluation (MIRACLE ICD) Trial Investigators. Combined cardiac resynchronization and implantable cardioversion defibrillation in advanced chronic heart failure: the MIRACLE ICD Trial. JAMA 2003;289:2685–94.

42. Cazeau S, Leclercq C, Lavergne T, et al., Multisite Stimulation in Cardiomyopathies (MUSTIC) Study Investigators. Effects of multisite biventricular pacing in patients with heart failure and intraventricular conduction delay. N Engl J Med 2001;344:873–80.

43. Abraham WT, Fisher WG, Smith AL, et al., MIRACLE Study Group. Cardiac resynchronization in chronic heart failure. N Engl J Med 2002;346:1845–53.

44. Gold MR, Van Veldhuisen DJ, Hauptman PJ, et al. Vagus nerve stimulation for the treatment of heart failure: the INOVATE-HF Trial. J Am Coll Cardiol 2016;68:149–58.

45. Middlekauff HR. Making the case for skeletal myopathy as the major limitation of exercise capacity in heart failure. Circ Heart Fail 2010;3:537–46.

46. Roul G, Germain P, Bareiss P. Does the 6-minute walk test predict the prognosis in patients with NYHA class II or III chronic heart failure? Am Heart J 1998;136:449–57.

47. Hülsmann M, Berger R, Sturm B, et al. Prediction of outcome by neurohumoral activation, the six-minute walk test and the Minnesota Living with Heart Failure Questionnaire in an outpatient cohort with congestive heart failure. Eur Heart J 2002;23:886–91.

48. Dunlay SM, Manemann SM, Chamberlain AM, et al. Activities of daily living and outcomes in heart failure. Circ Heart Fail 2015;8:261–7.

49. Auricchio A, Stellbrink C, Butter C, et al., Guidant Heart Failure Research Group. Clinical efficacy of cardiac resynchronization therapy using left ventricular pacing in heart failure patients stratified by severity of ventricular conduction delay. J Am Coll Cardiol 2003;42:2109–16.

50. Toblli JE, Lombraña A, Duarte P, et al. Intravenous iron reduces NT-pro-brain natriuretic peptide in anemic patients with chronic heart failure and renal insufficiency. J Am Coll Cardiol 2007;50:1657–65.

51. Tate CW III, Robertson AD, Zolty R, et al. Quality of life and prognosis in heart failure: results of the Beta-Blocker Evaluation of Survival Trial (BEST). J Card Fail 2007;13:732–7.

52. Flynn KE, Piña IL, Whellan DJ, et al., HF-ACTION Investigators. Effects of exercise training on health status in patients with chronic heart failure: HF-ACTION randomized controlled trial. JAMA 2009;301:1451–9.

53. O'Connor CM, Whellan DJ, Lee KL, et al., HF-ACTION Investigators. Efficacy and safety of exercise training in patients with chronic heart failure: HF-ACTION randomized controlled trial. JAMA 2009;301:1439–50.

54. Anker SD, Comin Colet J, Filippatos G, et al., FAIR-HF Trial Investigators. Ferric carboxymaltose in patients with heart failure and iron deficiency. N Engl J Med 2009;361:2436–48.

55. Ponikowski P, van Veldhuisen DJ, Comin-Colet J, et al., 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 2015;36:657–68.

56. Costanzo MR, Ivanhoe RJ, Kao A, et al. Prospective evaluation of elastic restraint to lessen the effects of heart failure (PEERLESS-HF) trial. J Card Fail 2012;18:446–58.

57. Moss AJ, Hall WJ, Cannom DS, et al., MADIT-CRT Trial Investigators. Cardiac-resynchronization therapy for the prevention of heart-failure events. N Engl J Med 2009;361:1329–38.

58. Veazie PJ, Noyes K, Li Q, et al. Cardiac resynchronization and quality of life in patients

with minimally symptomatic heart failure. J Am Coll Cardiol 2012;60:1940–4.

59. Pitt B, Pfeffer MA, Assmann SF, et al., TOPCAT Investigators. Spironolactone for heart failure with preserved ejection fraction. N Engl J Med 2014; 370:1383–92.

60. Lewis EF, Kim HY, Claggett B, et al., TOPCAT Investigators. Impact of spironolactone on longitudinal changes in health-related quality of life in the Treatment of Preserved Cardiac Function Heart Failure With an Aldosterone Antagonist trial. Circ Heart Fail 2016;9:e001937.

61. Pfeffer MA, Claggett B, Assmann SF, et al. Regional variation in patients and outcomes in the Treatment of Preserved Cardiac Function Heart Failure With an Aldosterone Antagonist (TOPCAT) trial. Circulation 2015;131:34–42.

62. Solomon SD, Clagget B, Lewis EF, et al., TOPCAT Investigators. Influence of ejection fraction on outcomes and efficacy of spironolactone in patients with heart failure with preserved ejection fraction. Eur Heart J 2016;37:455–62.

63. Shah AM, Claggett B, Sweitzer NK, et al. Prognostic importance of impaired systolic function in heart failure with preserved ejection fraction and the impact of spironolactone. Circulation 2015;132:402–14.

64. Shah AM, Claggett B, Sweitzer NK, et al. Prognostic importance of changes in cardiac structure and function in heart failure with preserved ejection fraction and the impact of spironolactone. Circ Heart Fail 2015;8:1052–8.

65. Ypenburg C, van Bommel RJ, Borleffs CJ, et al. Long-term prognosis after cardiac resynchronization therapy is related to the extent of left ventricular reverse remodeling at midterm follow-up. J Am Coll Cardiol 2009;53:483–90.

66. Suzuki H, Kusuyama T, Omori Y, et al. Inhibitory effect of candesartan cilexetil on left ventricular remodeling after myocardial infarction. Int Heart J 2006;47:715–25.

67. Cicoira M, Zanolla L, Rossi A, et al. Long-term, dose-dependent effects of spironolactone on left ventricular function and exercise tolerance in patients with chronic heart failure. J Am Coll Cardiol 2002;40:304–10.

68. Konstam MA, Kramer DG, Patel AR, et al. Left ventricular remodeling in heart failure: current concepts in clinical significance and assessment. J Am Coll Cardiol Img 2011;4:98–108.

69. McMurray JJV, Packer M, Desai AS, et al., PARADIGM-HF Investigators and Committees. Angiotensin-neprilysin inhibition versus enalapril in heart failure. N Engl J Med 2014;371:993–1004.

70. Lippi G, Sanchis-Gomar F. Monitoring B-type natriuretic peptide in patients undergoing therapy with neprilysin inhibitors. An emerging challenge? Int J Cardiol 2016;219:111–4.

71. Semenov AG, Katrukha AG. Different susceptibility of B-type natriuretic peptide (BNP) and BNP precursor (proBNP) to cleavage by neprilysin: the N-terminal part does matter. Clin Chem 2016; 62:617–22.

72. Bhatt DL, Mehta C. Adaptive designs for clinical trials. N Engl J Med 2016;375:65–74.

73. Saville BR, Connor JT, Ayers GD, et al. The utility of Bayesian predictive probabilities for interim monitoring of clinical trials. Clin Trials 2014;11:485–93.

74. Berry SM, Connor JT, Lewis RJ. The platform trial: an efficient strategy for evaluating multiple treatments. JAMA 2015;313:1619–20.

75. Saville BR, Berry SM. Efficiencies of platform clinical trials: a vision of the future. Clin Trials 2016;13:358–66.

76. Wick J, Berry SM, Yeh HW, et al. A novel evaluation of optimality for randomized controlled trials. J Biopharm Stat 2016 Jun 13 [E-pub ahead of print], http://dx.doi.org/10.1080/10543406.2016.1198367.

77. McKelvie RS, Yusuf S, Pericak D, et al. Comparison of candesartan, enalapril, and their combination in congestive heart failure. Circulation 1999;100:1056–64.

78. Conraads VM, Beckers P, Vaes J, et al. Combined endurance/resistance training reduces NT-proBNP levels in patients with chronic heart failure. Eur Heart J 2004;25:1797–805.

79. Krum H, Ashton E, Reid C, et al. Double-blind, randomized, placebo-controlled study of high-dose HMG CoA reductase inhibitor therapy on ventricular remodeling, pro-inflammatory cytokines and neurohormonal parameters in patients with chronic systolic heart failure. J Card Fail 2007;13:1–7.

80. Pascual-Figal DA, Domingo M, Casas T, et al. Usefulness of clinical and NT-proBNP monitoring for prognostic guidance in destabilized heart failure outpatients. Eur Heart J 2008;29:1011–8.

81. Bielecka-Dabrowa A, Goch JH, Mikhailidis DP, et al. The influence of atorvastatin on parameters of inflammation and function of the left ventricle in patients with dilated cardiomyopathy. Med Sci Monit 2009;15:MS12–23.

82. Zannad F, De Ferrari GM, Tuinenburg AE, et al. Chronic vagal stimulation for the treatment of low ejection fraction heart failure: results of the NEural Cardiac TherApy foR Heart Failure (NECTAR-HF) randomized controlled trial. Eur Heart J 2014; 36:425–33.

83. Ordu S, Yildiz BS, Alihanoglu YI, et al. Effects of ivabradine therapy on heart failure biomarkers. Cardiol J 2015;22:501–9.

84. Gheorghiade M, Greene SJ, Butler J, et al., SOCRATES-REDUCED Investigators and Coordinators. Effect of vericiguat, a soluble guanylate cyclase stimulator, on natriuretic peptide levels in patients with worsening chronic heart failure and reduced ejection fraction: the SOCRATES-REDUCED Randomized Trial. JAMA 2015;31:2251–62.

85. Lipkin DP, Canepa-Anson R, Stephens M, et al. Factors determining symptoms in heart failure: comparison of fast and slow exercise tests. Br Heart J 1986;55:439–45.

86. Guazzi M, Dickstein K, Vicenzi M, et al. Six-minute walk test and cardiopulmonary exercise testing in patients with chronic heart failure: a comparative analysis on clinical and prognostic insights. Circ Heart Fail 2009;2:549–55.

87. Täger T, Hanholz W, Cebola R, et al. Minimal important difference for 6-minute walk test distances among patients with chronic heart failure. Int J Cardiol 2014;176:94–8.

88. Ingle L, Cleland JG, Clark AL. The relation between repeated 6-minute walk test performance and outcome in patients with chronic heart failure. Ann Phys Rehabil Med 2014;57: 244–53.

KEY WORDS heart failure trials, natriuretic peptides, surrogate endpoints