European Journal of Heart Failure (2016) **18**, 482–489
doi:10.1002/ejhf.516

REVIEW

Traditional and new composite endpoints in heart failure clinical trials: facilitating comprehensive efficacy assessments and improving trial efficiency

Stefan D. Anker[1]*, Stefan Schroeder[2], Dan Atar[3], Jeroen J. Bax[4], Claudio Ceconi[5], Martin R. Cowie[6], Adam Crisp[7], Fabienne Dominjon[8], Ian Ford[9], Hossein-Ardeschir Ghofrani[10,11,12], Savion Gropper[13], Gerhard Hindricks[14], Mark A. Hlatky[15], Richard Holcomb[16], Narimon Honarpour[17], J. Wouter Jukema[4], Albert M. Kim[18], Michael Kunz[2], Martin Lefkowitz[19], Chantal Le Floch[8], Ulf Landmesser[20], Theresa A. McDonagh[21], John J. McMurray[22], Bela Merkely[23], Milton Packer[24], Krishna Prasad[25], James Revkin[18], Giuseppe M.C. Rosano[26,27], Ransi Somaratne[17], Wendy Gattis Stough[28], Adriaan A. Voors[29], and Frank Ruschitzka[30]

[1]University Medical Center Gottingen, Gottingen, Germany; [2]Bayer Healthcare, Berlin, Germany; [3]Department of Cardiology B, Oslo University Hospital Ulleval and University of Oslo, Norway; [4]Department of Cardiology, Leiden University Medical Center, Leiden, The Netherlands; [5]Unit of Cardiology, Hospital of Desenzano del Garda, Desenzano del Garda, Italy; [6]National Heart and Lung Institute, Imperial College London, London, UK; [7]GlaxoSmithKline, Middlesex, UK; [8]Servier International, Paris, France; [9]Robertson Centre for Biostatistics, University of Glasgow, Glasgow, UK; [10]University of Giessen and Marburg Lung Center, UGMLC [member of the German Center for Lung Research (DZL)], Giessen, Germany; [11]Kerckhoff-Klinik Bad Nauheim, Germany; [12]Imperial College London, UK; [13]Boehringer-Ingelheim GmBH & Co. KG, Ingelheim, Germany; [14]Department of Electrophysiology, Heart Center, University of Leipzig, Leipzig, Germany; [15]Stanford University School of Medicine, Stanford, CA, USA; [16]Independent Biostatistician, Minneapolis, MN, USA; [17]Amgen, Inc., Thousand Oaks, CA, USA; [18]Pfizer, Inc., Cambridge, MA, USA; [19]Novartis Pharmaceuticals, East Hanover, NJ, USA; [20]Department of Cardiology, Charité Universitätsmedizin Berlin, Berlin, Germany; [21]King's College Hospital, London, UK; [22]BHF Cardiovascular Research Centre, University of Glasgow, Glasgow, UK; [23]Semmelweis University Heart and Vascular Center, Budapest, Hungary; [24]Department of Clinical Sciences, University of Texas Southwestern Medical Center, Dallas, TX, USA; [25]United Kingdom Medicines and Healthcare Products Regulatory Agency, London, UK; [26]IRCCS San Raffaele Hospital Roma, Rome, Italy; [27]Cardiovascular and Cell Sciences Institute, St. George's University of London, London, UK; [28]Campbell University College of Pharmacy and Health Sciences, NC, USA; [29]University of Groningen, Groningen, The Netherlands; and [30]Department of Cardiology, Heart Failure Clinic and Transplantation, University Heart Center Zurich, Zurich, Switzerland

Received 1 October 2015; revised 15 December 2015; accepted 17 January 2016; online publish-ahead-of-print 12 April 2016

Composite endpoints are commonly used as the primary measure of efficacy in heart failure clinical trials to assess the overall treatment effect and to increase the efficiency of trials. Clinical trials still must enrol large numbers of patients to accrue a sufficient number of outcome events and have adequate power to draw conclusions about the efficacy and safety of new treatments for heart failure. Additionally, the societal and health system perspectives on heart failure have raised interest in ascertaining the effects of therapy on outcomes such as repeat hospitalization and the patient's burden of disease. Thus, novel methods for using composite endpoints in clinical trials (e.g. clinical status composite endpoints, recurrent event analyses) are being applied in current and planned trials. Endpoints that measure functional status or reflect the patient experience are important but used cautiously because heart failure treatments may improve function yet have adverse effects on mortality. This paper discusses the use of traditional and new composite endpoints, identifies qualities of robust composites, and outlines opportunities for future research.

Keywords Heart failure • Clinical trial • Endpoint determination • Surrogate endpoints

*Corresponding author: Division of Innovative Clinical Trials, Department of Cardiology & Pneumology, University Medical Center Göttingen (UMG), Robert-Koch-Str. 40, D-37075 Göttingen, Germany. Tel: +49 551 39 2091, Fax: +49 551 39 20911, Email: s.anker@cachexia.de

Introduction

Composite endpoints are increasingly used as primary efficacy measures in heart failure clinical trials (Supplementary material online, *Table S1*)[1] to provide a comprehensive picture of the treatment effect, and to improve trial efficiency by increasing the event rate and reducing the required sample size. They have advantages and disadvantages (*Table 1*),[1–4] but composite endpoints are generally accepted by academics, clinicians, and regulators when the components are well-defined, specific to the key objective of interest, and broadly congruent in regards to treatment effect. They are problematic when the overall effect suggests no benefit, or even harm, in one or more components.

The adoption of evidence-based therapies has reduced event rates for well-established composites (e.g. all-cause mortality plus cardiovascular hospitalization; cardiovascular death plus heart failure hospitalization). Thus, large studies with long follow-up are still needed in modern trials to accrue a sufficient number of events for adequate statistical power. These studies require substantial investment from funders (e.g. public or industry sources) who must decide whether to commit significant funds when there is a real risk that the therapy will be ineffective. Industry investment in cardiovascular drug development has been decreasing, perhaps in part because of these reasons,[5,6] and cardiovascular disease is no longer one of the top 10 therapeutic areas for research and development.[7]

It is a matter of debate whether the composite of cardiovascular mortality or heart failure hospitalization is the most meaningful and clinically relevant endpoint. New endpoints (e.g. novel clinical composites, functional measures, or patient-reported outcomes) or analytical methodologies (e.g. recurrent event analyses, responder analyses) might serve the dual purpose of more accurately reflecting the modern heart failure patient's disease burden and improving trial efficiency. However, confirmation of their validity is needed before they can achieve widespread acceptance.

The Cardiovascular Round Table (CRT) of the European Society of Cardiology (ESC) convened a 2-day workshop to explore how existing and innovative composite endpoints can be leveraged to advance the conduct of heart failure clinical trials and, ultimately, patient care. Workshop participants identified five qualities that should characterize composite endpoints in heart failure clinical trials (*Table 2*). This paper summarizes the key insights and discussions, suggests approaches for using composite endpoints, and identifies knowledge gaps that need to be addressed by further research.

Overview of composite endpoints in heart failure trials

Fatal and non-fatal composite clinical outcomes

All-cause assessments of fatal and non-fatal outcomes (e.g. all-cause mortality plus all-cause hospitalization) reflect an intervention's net benefit. Since a single intervention is unlikely to reduce all modes of death or causes of hospitalization, a significant reduction in an all-cause composite endpoint can be interpreted to indicate that the intervention reduced the major causes of death or hospitalization (usually cardiovascular in heart failure trials) without significant adverse effects.[1]

Estimated treatment effects may be diluted if a substantial proportion of events are not influenced by the treatment.[8] A greater number of events will increase statistical power only if those additional events are potentially modifiable by the intervention. The addition of outcomes that are not influenced by the treatment will reduce the measured treatment effect and the study power. Thus, composite endpoints should only include components that are relevant to the population being studied and have biological plausibility to support an expected treatment benefit on each component. A significant risk reduction in a composite endpoint does not necessarily imply that the risk reduction is equal across each component. Statistical methods to test for heterogeneity of treatment effect across individual components of a composite endpoint have been described but are likely to be underpowered.[9,10] Typically, reporting the individual components and testing the treatment effect on each is sufficient.

Similar logic applies to all-cause vs. cause-specific endpoints. The Candesartan in Heart Failure Assessment of Reduction in Mortality and Morbidity (CHARM) trials, which included a broad spectrum of heart failure patients, used all-cause mortality as the primary endpoint for the overall programme.[11] Out of 1831 deaths, 371 were non-cardiovascular and unlikely to be influenced by cardiovascular therapy. While the unadjusted hazard ratio (HR) for the effect of candesartan on all-cause mortality was not statistically significant [HR 0.91, 95% confidence interval (CI) 0.83–1.00, $P = 0.055$], the effect on cardiovascular mortality (a cause-specific endpoint) was significant (HR 0.88, 95% CI 0.79–0.97, $P = 0.012$).[11]

Owing in part to this experience, heart failure trials are often now designed with cause-specific composite endpoints (e.g. cardiovascular mortality or heart failure hospitalization). Cause-specific endpoints reflect a more precise evaluation of an intervention's expected effect, since it is unlikely that a cardiovascular drug will reduce non-cardiovascular causes of death (e.g. cancer, accidents). When cause-specific primary endpoints are used, all-cause mortality should still be evaluated as a secondary or safety endpoint to ensure that survival is not adversely affected by another pathway. The effect on all-cause mortality should be directionally similar to the effect on cardiovascular mortality (if the majority of deaths are disease related), even if the effect on all-cause mortality is not statistically significant. If cardiovascular mortality is significantly reduced and all-cause mortality is shifted towards the null, further analyses are needed to explain the findings and explore whether some component of non-cardiovascular mortality was increased in the treated vs. the control arm. Similarly, discordant results in individual components may be concerning and warrant further investigation, even if risk is reduced in the overall composite. Such analyses are exploratory and not likely to be definitive (depending on the number of events), but they can inform further research strategies to clarify the non-cardiovascular or discordant effects.

Limitations of cause-specific endpoints include the need for specific event definitions and endpoint adjudication. The precision of adjudication depends on complete medical records, which may

Table 1 Advantages and disadvantages of composite endpoints

Advantages	Disadvantages
Reduces sample size and improves trial efficiency if each component is modifiable by treatment Power depends on number of events not number of patients	May reduce power if some components of the composite are unaffected by treatment
May allow adequate accrual of events over shorter follow-up	Length of follow-up may be inadequate to characterize safety profile of a therapy intended for chronic use
Captures multiple aspects of treatment effect (not limited to survival)	Overall treatment effect may be driven by components of lesser importance (i.e. the least serious events usually occur with the greatest frequency)
Avoids competing risk	Treatment effects on different components may be directionally different.
Reduces need for multiple comparisons and allocation of type 1 error across different endpoints	In time-to-first event analyses, the first event may not be the most serious.
	Individual components still need to be reported separately (usually as secondary endpoints)
	Insufficient power to draw conclusions about treatment effects for all components of the composite
	Uncertainty regarding worrisome trends in components of the composite.

Table 2 Characteristics of robust composite endpoints

A composite endpoint should:

1. Provide reliable and precise estimates of efficacy and safety
2. Be clinically meaningful or relevant to physicians, patients, and care providers in terms of characterizing disease progression, stabilization, or reversal
3. Meaningfully characterize the burden of disease for patients
4. Yield information that could be used in conjunction with other data to determine societal valuation of new therapies or interventions
5. Improve the efficiency of clinical trials while maintaining high validity and quality

be limited for out-of-hospital deaths or some regions in global trials. External factors (e.g. geographical standards of care, threshold for admission, reimbursement pressures, availability of outpatient treatment, and medical or dietary non-adherence) may influence cause-specific hospitalization, although proper randomization should minimize the impact of these factors. These factors may have greater impact if a majority of patients are enrolled in regions with large differences in standards of care.

Reporting the individual components of composite endpoints is important to examine whether or not they are concordant, but non-fatal endpoints should not be analysed independently because of the problem of competing risks.[12] Composite endpoints can solve this dilemma if they include both fatal and non-fatal events. However, analysing fatal and non-fatal events in a composite endpoint can be problematic because these events differ in their importance. The least serious events (i.e. hospitalizations in heart failure trials, non-fatal myocardial infarctions in acute coronary syndrome trials) usually occur earlier than more serious events (i.e. death). Time-to-event analysis focuses on the first event. Thus, composite endpoints are often driven by the least serious component,[13] which decreases the relevance (not confidence) of any finding. For example, in the Systolic Heart Failure Treatment with the I_f Inhibitor Ivabradine Trial (SHIFT) study, the primary endpoint was a composite of cardiovascular death or hospital admission for worsening heart failure.[14] Ivabradine reduced the risk of the primary endpoint compared with placebo in patients with a heart rate ≥70 b.p.m. (HR 0.82, 95% CI 0.75–0.90, $P < 0.0001$), mainly driven by a reduction in heart failure hospitalizations (16% vs. 21%, HR 0.74, 95% CI 0.66–0.83, $P < 0.0001$) with no effect on cardiovascular mortality (14% vs. 15%, HR 0.91, 95% CI 0.80–1.03, $P = 0.13$).[14] The European Medicines Agency (EMA) approved ivabradine, but only for patients with heart rates >75 b.p.m. where a possible nominal reduction in overall mortality was observed.[15] Food and Drug Administration (FDA) approval was recently granted for a reduction in heart failure hospitalization in patients with a heart rate ≥70 b.p.m. Based on results from the Valsartan Heart Failure Trial (Val-HeFT), valsartan was approved by the FDA to reduce hospitalization for heart failure, with no indication for improvement in mortality because it reduced only one of the two primary endpoints: (i) all-cause mortality (relative risk 1.02, 95% CI 0.88–1.18, $P = 0.80$) and (ii) the composite of all-cause mortality or cardiac arrest with resuscitation, hospitalization for heart failure, or administration of intravenous inotropes or vasodilators for ≥4 h without hospitalization (relative risk 0.87, 95% CI 0.77–0.97, $P = 0.009$). The beneficial effect of valsartan on the latter was driven by a reduction in heart failure hospitalization.[16] Statistical methods to weight outcomes according to severity have been proposed in heart failure and other disease states.[17–22] However, these approaches are limited by lack of consensus on the relative weighting of events and inconsistency across studies.

These examples underscore the importance of limiting composites to include events that are clinically meaningful and considered to be modifiable. It is highly recommended that regulatory agencies be involved early in the process of constructing composite endpoints for use in pivotal trials.

Clinical status composite endpoints

In chronic conditions such as heart failure, mortality is not the only meaningful efficacy measure, since a patient may be alive but have a poor clinical status, functional capacity, or quality of life.[23] A treatment may be worthwhile and considered valuable by patients

when it improves their clinical status even if it does not prolong their survival.[24,25] A clinical status composite endpoint has been developed[26] and used[27–30] in heart failure trials (with adaptations appropriate for the study population). Patients are categorized as improved (moderate or marked improvement in clinical status at all planned assessments without hospitalization for heart failure or death); unchanged (modest improvement or worsening in clinical status); or worsened (moderate or marked worsening of clinical status at any planned assessment, hospitalization for heart failure requiring intravenous or mechanical interventions, or death). The distribution of responses can be compared between treatment groups without assigning ranks or worse scores to the individual components.[26] This method has the advantage of describing the patient's clinical course and response to treatment, but more experience is needed to validate that it produces a reliable, clinically meaningful, and unbiased estimate of treatment effect. Analyses can determine if a statistically significant difference exists in the distribution of patients who improved, worsened, or were unchanged, but the magnitude or clinical relevance of the effect can be challenging to interpret. Achieving consensus regarding what constitutes clinically meaningful degrees of improvement and worsening (recognizing that the criteria will differ to some extent by study population) is one challenge of implementing this endpoint. Variation in standards for hospitalizations across geographic regions may also be problematic, but capturing moderate to marked worsening in symptoms or functional status might minimize the problem of geographic variation in standards of care. Finally, experience with this endpoint suggests that it is most appropriate for relatively short-term trials (<9 months) because of the challenges related to assessing changes in a patient's clinical improvement over lengthy time periods (i.e. patient recall) and because more clinical events accrue during long-term follow-up which outweigh the clinical status improvement component of the composite.

Functional endpoints

Most clinicians agree that functional impairment is a primary concern for patients with heart failure. For some patients, improved functional status or quality of life is of greater importance than longevity.[2,25] Patients are surviving longer with heart failure because of therapeutic advances,[31] so assessing functional status may be of even greater importance in this era of improved survival. The primary goal of considering novel composite endpoints that include functional status or patient-reported symptom measures is to identify treatments that improve important aspects of patient well-being beyond survival. However, few, if any, cardiovascular drugs have been approved on the basis of improved functional status alone. Regulators have been cautious about functional status endpoints in cardiovascular clinical trials because of agents that improved exercise tolerance but increased mortality in large trials (e.g. flosequinan).[32] The problem of defining a clinically meaningful change in exercise time (or other measure of functional status) also contributes to the uncertainty about the value of this endpoint.

A clinically meaningful improvement in functional status may lead to approval of a new therapeutic agent, provided that an adequate margin of safety can be assured.[1] A stringent margin for excluding an adverse effect on mortality would probably be required in a trial using a functional primary endpoint. Thus, efficiencies gained by using novel composites may be offset by the need to demonstrate safety, since studies would still need to be large and long enough to rule out an increased risk of mortality (at some threshold level acceptable to regulatory agencies) or provide reasonable assurance of a neutral effect on mortality.

The persisting question is what endpoint(s) (other than death or hospitalization) might be clinically relevant and scientifically valid, while also increasing the efficiency of drug development and clinical trial conduct. The 6-min walk test distance has been the primary endpoint in most registration trials of pulmonary artery hypertension,[33,34] accepted by both the EMA and FDA, but it has not been recognized as adequate in heart failure. Other functional measures such as peak VO_2 might be candidate endpoints for heart failure trials, but they are unphysiological, can be influenced by patient motivation and skeletal muscle function, and may not be consistently reproducible, which limits their use as an endpoint in heart failure and other (e.g. pulmonary hypertension) trials. Health-related quality of life as measured by instruments such as the Minnesota Living With Heart Failure (MLWHF) or the Kansas City Cardiomyopathy Questionnaire (KCCQ) might be considered as endpoints and have been used in many studies, but the FDA has specific standards that must be met when patient-reported outcomes are used to support labelling claims.[35,36] While health-related quality of life assessments are clinically relevant and can be informative, the methodological problems of using such scores in efficacy assessments (e.g. potential for bias particularly in unblinded studies, procedures for handling missing data) are well known.[35] Whether regulatory agencies would accept the MLWHF or KCCQ scores as supportive evidence of efficacy remains to be seen and would probably need to be considered on an individual trial basis.

Finally, another important consideration is the health technology assessment of new drugs after regulatory approval. Even if an endpoint such as 6-min walk distance were accepted by regulators, it is uncertain whether payers would view it as a worthwhile endpoint. Research would need to validate the level of increased exercise tolerance that was cost-effective and had a societal benefit.

Role of adaptive licensing

In 2007, the EMA described its openness to innovative drug development approaches.[37] One such initiative is the Pilot Project on Adaptive Licensing.[38] Adaptive licensing involves an authorized limited indication followed by 'iterative phases of evidence gathering and progressive licensing adaptations concerning both the authorized indication and the potential further therapeutic uses of the drug'.[38] With adaptive licensing, a drug could be approved based on improvement in a well-defined functional endpoint within a rigorously conducted clinical trial. After licensing, post-authorization efficacy and safety studies would be required.[38–40] This process fulfils the goal of accelerating patient access to new drugs, while providing a mechanism for collecting safety data. Functional status

or patient-reported symptom endpoints could have a role in adaptive licensing, but past experiences in heart failure where initially promising drugs have later been found harmful (e.g. ibopamine, flosequinan) emphasize the need to pursue this approach cautiously. Many issues have been identified with this approach. First, it is uncertain if a relatively short-term trial assessing a functional endpoint will accrue a sufficient number of events to provide early estimates of safety prior to granting an adaptive licence. Additionally, the safety margin for excluding excess risk needs to be defined and achieving harmonization among regulatory bodies could be problematic. Presumably, the acceptable safety margin could vary by patient population, severity of illness, or the pre-test probability of risk in the context of the mechanism of action of the drug or device; thus, safety estimates should account for this variation. On the other hand, the adaptive approval process aims to accrue pharmacovigilance data in the early phase of marketing through registries, thereby monitoring the real-life use and adverse event rates of the drugs approved with this regulatory pathway.[41] Conducting randomized clinical trials after a drug is marketed and available to patients is more difficult than pre-approval from the standpoint of subject recruitment and retention. All patients receiving the drug could be followed in a registry for safety, but the results of observational studies are less reliable because of the potential for bias. The current system is not optimally designed to allow regulators to enforce withdrawal of adaptive licences if the sponsor does not uphold the requirements of the licence (e.g. follow-up trials are not completed, a concerning safety signal emerges but falls short of crossing pre-defined margins of licence withdrawal, or safety trials are poorly designed), particularly in the European Union where decisions to withdraw drugs are made by each member state. One approach would be to apply adaptive licensing only to severely ill patients where the balance of risks and benefits might be more favourable. However, once marketed, restrictions on use according to patient severity will be difficult to enforce.

Statistical considerations: new analytical methods

Recurrent event analyses

Heart failure is a chronic disease, typically characterized by repeat hospitalizations as a patient's condition progressively worsens. Several factors have contributed to an interest in analysing recurrent events in heart failure trials.[42] Hospitalization is the major contributor to the overall cost of heart failure care,[43] which has led to targeted interventions to reduce readmissions.[44,45] Therefore, data describing an intervention's effect on recurrent events is highly clinically relevant. Importantly, restricting analyses to first events incompletely represents the patient's overall burden of disease, since first events account for only half of the total number of heart failure hospitalizations in major clinical trials.[46–49] Several approaches to recurrent event analyses have been tested using major heart failure trial data sets.[46–51] All have limitations, as complex assumptions are made in determining the study sample size and in the statistical modelling.[52] One concern is overestimation of the treatment effect,[48] and another is how best to manage the competing risk of death. Either the joint frailty model or the negative binomial distribution (if death rates are low) have been used,[48,50] but none of the methods has been well validated. Regulatory acceptance of the methodology is a critical aspect, and the agencies appear to be supportive of an analysis strategy based on recurrent events. The primary endpoint of the PARAGON-HF study (Prospective Comparison of ARNI with ARB Global Outcomes in Heart Failure With Preserved Ejection Fraction, NCT01920711) is the cumulative number of primary composite events of cardiovascular death and total (first and recurrent) heart failure hospitalizations.[53] PARAGON-HF was also designed to have reasonable power on a standard time-to-first event analysis for sensitivity comparisons. This approach allows regulators and investigators to gain experience with a recurrent events method against a background of traditional time-to-first event analysis.

Responder analyses

Responder analyses are used to assess the clinical relevance of an observed treatment effect.[36,54,55] From the EMA perspective, responder analyses have been used in studies of obesity (e.g. comparison of proportion of patients with $\geq$10% weight loss), acute stroke (e.g. comparison of proportion of survivors who regain functional independence), or depression (e.g. comparison of proportion of patients achieving a threshold change in symptom scores). The challenge with responder analyses, especially in heart failure trials, is defining, validating, and achieving consensus on the clinically meaningful change in a functional measure or symptom score.[35,56–58] In the context of a heart failure trial, a responder analysis could be based on a surrogate variable that is expected to be related to reliable clinical improvement or clinical outcome. Historically, measures of functional status on which responder analyses might be based (e.g. exercise time, LV size, EF, or haemodynamics) inconsistently correlate with clinical outcome. Scales to measure symptoms, dyspnoea, functional status, and health-related quality of life are alternatives to surrogates and have been used across clinical trials. Interpretation of treatment effect is difficult since effect size for the same treatment can differ when measured on different scales, and definition of a clinically important change can be arbitrary and that attenuation of decline, i.e. no change, may itself be an indicator of benefit. Importantly, study power can be substantially reduced in a responder analysis when a continuous outcome measure is dichotomized.[59]

Responder analyses might be useful in phase I and II trials to help inform phase III designs and patient selection criteria; however, many well-known examples exist where favourable phase II results did not translate into improved outcome in phase III.[60] Therefore, cautious interpretation of phase II results used for this purpose is advised. On the other hand, if efficacy is demonstrated using traditional methods, responder analyses can be performed secondarily to characterize the clinical relevance of the effect and to evaluate the potential for hyper-responders (in cardiac resynchronization therapy trials often called super-responders), non-responders, or patients who experience harm. Signals of efficacy or harm observed in subgroups are far from definitive, but they can generate new hypotheses for testing in adequately

Table 3 Areas of uncertainty, and priorities for future research

Topic	Description
Component selection	1. Involve regulators early in discussion of which components will be included in a composite endpoint; may differ between acute and chronic heart failure trials 2. Explicitly state the rationale for using a composite in the study protocol
Clinical status composite endpoints	1. Which variables accurately reflect an unbiased assessment of a patient's clinical status and burden of disease? 2. How can clinical composite endpoints be translated clinically (in terms of magnitude of effect and clinical relevance)? 3. How can quality of extended life be assessed? 4. Propose a quantitative assessment of the concordance of components of composite endpoints
Recurrent event analysis	1. Methodology needs to be further studied and refined (e.g. how to deal with event clustering, ensure analysis is not driven by a small proportion of patients). 2. How can accuracy of assumptions and simulations used in modelling be confirmed? 3. Involve regulators in planning modelling techniques for recurrent event analyses 4. Perform sensitivity analyses using standard time-to-event analysis 5. How to interpret results if recurrent event analysis results differ substantially in magnitude or direction from time-to-first event analysis.

powered studies within the specific population that appeared to have the greatest treatment effect. Assessing heterogeneity of benefit or net benefit, balancing benefit and harm, using subgroups based on a risk score has been proposed as a more powerful approach to subgroup analysis.[61,62] Responder analyses are also of interest to support health technology assessment. However, it is critical to acknowledge the limitations of these analyses to avoid overinterpretation of the data.

Conclusion

The changing landscape for heart failure clinical trials has created an important opportunity to learn from past successes (and failures) and shape future approaches. The use of traditional composite endpoints has yielded many highly effective therapies, and the approach should not be abandoned. However, the overall health burden from heart failure in patients is broader than hospitalizations and death, and it will be necessary to conduct trials that reflect endpoints important to patients[35] (and payers). Further, the declining resources for conducting larger, longer, and more costly trials is a reality that cannot be ignored.[5] Composites that reflect both clinical status and traditional 'hard' events, new analytical methodologies to assess recurrent events, and renewed efforts to assess changes in functional status (without ignoring safety) are areas of both interest and uncertainty that will benefit from additional research (*Table 3*). Ongoing trials will provide more insight into the strengths and weaknesses of these approaches and inform future directions in heart failure research.

Acknowledgements

This paper was generated from discussions during a Cardiovascular Round Table (CRT) Workshop organized on 21–22 January 2015 by the European Society of Cardiology (ESC). The CRT is a strategic forum for high-level dialogues between industry and ESC leadership to identify and discuss key strategic issues for the future of cardiovascular health in Europe.

The opinions expressed in this paper are those of the authors and cannot be interpreted as the opinion of any of the organizations that employ the authors. The authors acknowledge the contribution of Caroline Boulton for participating in the discussion at the ESC CRT workshop from which this paper originated.

Supplementary Information

Additional Supporting Information may be found in the online version of this article:

Table S1. Composite endpoints in heart failure clinical trials.

Conflicts of interest: S.D.A. has received esearch grants from Vifor and Abbott Vascular; and personal fees from Vifor, Abbott Vascular, Bayer Pharma, Novartis, Lonestar Heart, Respicardia, ZS Pharma, Relypsa, Biotronik, Cardiorentis, and Servier. S.S. is an employee of Bayer Healthcare. D.A, has received personal fees from Astra-Zeneca, Bayer Healthcare, Boehringer-Ingelheim, BMS/Pfizer, Novartis, and Vifor. J.J.B. reports that the Department of Cardiology (Leiden University Medical Center, The Netherlands) has received unrestricted research grants from Medtronic, Biotronik, Boston Scientific, and Edwards Lifescience. M.R.C. has received research grants from ResMed, Boston Scientific, and Bayer; and personal fees from ResMed, Boston Scientific, Bayer, Servier, Novartis, St. Jude Medical, and Pfizer. A.C. is an employee of GlaxoSmithKline. F.D. and C.L.F. are employees of Servier. I.F. reports grants from Vifor, grants and personal fees from Servier and Biotronik, and personal fees from RESMED, outside the submitted work. H.-A.G, has received personal fees (board membership, consultancy, paid lectures, or travel expenses) from Actelion, Bayer, GSK, Novartis, Pfizer, Bellerophon Therapeutics, United Therapeutics; and research grants from Actelion, Bayer, Pfizer, and Novartis. S.G. is an employee of Boehringer-Ingelheim GmbH & Co.KG. G.H. has received research grants from St. Jude

Medical, Biotronik, and Boston Scientific. N.H. is an employee of Amgen. J.W.J. has received research grants from and CME speaker fees for meeting sponsored by Astellas, Astra Zeneca, Bayer, Biotronik, Boston Scientific, Daiichi Sankyo, Lilly, Genzyme, Medtronic, Merck Schering Plough, Pfizer, Orbus Neich, Novartis, Roche, Servier, Sanofi Aventis, The Netherlands Heart Foundation, Interuniversity Cardiology Institute of the Netherlands, and the European Community Framework KP7 Programme. A.M.K. is an employee of Pfizer. M.K. is an employee of Bayer Healthcare. M.L. is an employee of Novartis Pharmaceuticals Company. U.L. has received research grants from Pfizer and Roche; and honoraria (lecture and advisory) from Roche, MSD, Sanofi, and Amgen. T.A.M. has received personal fees (honoraria) from Novartis, Vifor, and ZS Pharma. B.M. has received personal fees (lectures and advisory board membership) from Orion Pharma. M.P. reports personal fees from Amgen, Daiichi Sankyo, Ferring, Cardiorentis, CardioKinetix, Novartis, Sanofi, and Takeda, outside the submitted work. J.R. reports personal fees from Pfizer, outside the submitted work. R.S. is an employee and shareholder of Amgen, Inc. W.G.S. has received personal fees from the European Society of Cardiology, Relypsa, Covis, and Stealth Peptides. A.A.V. reports research grants from Novartis Pharmaceuticals, Singulex, Sphingotec, Trevena, and Vifor; personal fees from Novartis Pharmaceuticals, Servier, Bayer, Boehringer Ingelheim, Cardio3Biosciences, Celladon, GSK, Merck/MSD, Singulex, Trevena, and Vifor; and non-financial support from Alere and Merck/MSD. The remaining authors have no conflicts to disclose.

References

1. Zannad F, Garcia AA, Anker SD, Armstrong PW, Calvo G, Cleland JG, Cohn JN, Dickstein K, Domanski MJ, Ekman I, Filippatos GS, Gheorghiade M, Hernandez AF, Jaarsma T, Koglin J, Konstam M, Kupfer S, Maggioni AP, Mebazaa A, Metra M, Nowack C, Pieske B, Pina IL, Pocock SJ, Ponikowski P, Rosano G, Ruilope LM, Ruschitzka F, Severin T, Solomon S, Stein K, Stockbridge NL, Stough WG, Swedberg K, Tavazzi L, Voors AA, Wasserman SM, Woehrle H, Zalewski A, McMurray JJ. Clinical outcome endpoints in heart failure trials: a European Society of Cardiology Heart Failure Association consensus document. *Eur J Heart Fail* 2013;**15**:1082–1094.
2. Cordoba G, Schwartz L, Woloshin S, Bae H, Gotzsche PC. Definition, reporting, and interpretation of composite outcomes in clinical trials: systematic review. *BMJ* 2010;**341**:c3920.
3. Freemantle N, Calvert M, Wood J, Eastaugh J, Griffin C. Composite outcomes in randomized trials: greater precision but with greater uncertainty? *JAMA* 2003;**289**:2554–2559.
4. European Network for Health Technology Assessment. Endpoints used for relative effectiveness assessment of pharmaceuticals: clinical endpoints. https://eunethta.fedimbo.belgium.be/sites/5026.fedimbo.belgium.be/files/Clinical%20endpoints.pdf (15 February 2015).
5. Jackson N, Atar D, Borentain M, Breithardt G, van Eickels M, Endres M, Fraass U, Friede T, Hannachi H, Janmohamed S, Kreuzer J, Landray M, Lautsch D, Le Floch C, Mol P, Naci H, Samani N, Svensson A, Thorstensen C, Tijssen J, Vandzhura V, Zalewski A, Kirchhof P. Improving clinical trials for cardiovascular diseases: a position paper from the Cardiovascular Round Table of the European Society of Cardiology. *Eur Heart J* 2016;**37**:747–754.
6. Roe MT, Mahaffey KW, Ezekowitz JA, Alexander JH, Goodman SG, Hernandez A, Temple T, Berdan L, Califf RM, Harrington RA, Peterson ED, Armstrong PW. The future of cardiovascular clinical research in North America and beyond-addressing challenges and leveraging opportunities through unique academic and grassroots collaborations. *Am Heart J* 2015;**169**:743–750.
7. Moses H III, Matheson DH, Cairns-Smith S, George BP, Palisch C, Dorsey ER. The anatomy of medical research: US and international comparisons. *JAMA* 2015;**313**:174–189.
8. Gomez G, Gomez-Mateu M, Dafni U. Informed choice of composite end points in cardiovascular trials. *Circ Cardiovasc Qual Outcomes* 2014;**7**:170–178.
9. Pogue J, Thabane L, Devereaux PJ, Yusuf S. Testing for heterogeneity among the components of a binary composite outcome in a clinical trial. *BMC Med Res Methodol* 2010;**10**:49.
10. Pogue J, Devereaux PJ, Thabane L, Yusuf S. Designing and analyzing clinical trials with composite outcomes: consideration of possible treatment differences between the individual outcomes. *PLoS One* 2012;**7**:e34785.
11. Pfeffer MA, Swedberg K, Granger CB, Held P, McMurray JJ, Michelson EL, Olofsson B, Ostergren J, Yusuf S, Pocock S, CHARM Investigators and Committees. Effects of candesartan on mortality and morbidity in patients with chronic heart failure: the CHARM-Overall programme. *Lancet* 2003;**362**:759–766.
12. Wolbers M, Koller MT, Stel VS, Schaer B, Jager KJ, Leffondre K, Heinze G. Competing risks analyses: objectives and approaches. *Eur Heart J* 2014;**35**:2936–2941.
13. Ferreira-Gonzalez I, Busse JW, Heels-Ansdell D, Montori VM, Akl EA, Bryant DM, onso-Coello P, Alonso J, Worster A, Upadhye S, Jaeschke R, Schunemann HJ, Permanyer-Miralda G, Pacheco-Huergo V, Domingo-Salvany A, Wu P, Mills EJ, Guyatt GH. Problems with use of composite end points in cardiovascular trials: systematic review of randomised controlled trials. *BMJ* 2007;**334**:786.
14. Swedberg K, Komajda M, Bohm M, Borer JS, Ford I, Dubost-Brama A, Lerebours G, Tavazzi L. Ivabradine and outcomes in chronic heart failure (SHIFT): a randomised placebo-controlled study. *Lancet* 2010;**376**:875–885.
15. Bohm M, Borer J, Ford I, Gonzalez-Juanatey JR, Komajda M, Lopez-Sendon J, Reil JC, Swedberg K, Tavazzi L. Heart rate at baseline influences the effect of ivabradine on cardiovascular outcomes in chronic heart failure: analysis from the SHIFT study. *Clin Res Cardiol* 2013;**102**:11–22.
16. Cohn JN, Tognoni G, Valsartan Heart Failure Trial Investigators. A randomized trial of the angiotensin-receptor blocker valsartan in chronic heart failure. *N Engl J Med* 2001;**345**:1667–1675.
17. Armstrong PW, Westerhout CM, Van de Werf F, Califf RM, Welsh RC, Wilcox RG, Bakal JA. Refining clinical trial composite outcomes: an application to the Assessment of the Safety and Efficacy of a New Thrombolytic-3 (ASSENT-3) trial. *Am Heart J* 2011;**161**:848–854.
18. Buyse M. Generalized pairwise comparisons of prioritized outcomes in the two-sample problem. *Stat Med* 2010;**29**:3245–3257.
19. Pocock SJ, Ariti CA, Collier TJ, Wang D. The win ratio: a new approach to the analysis of composite endpoints in clinical trials based on clinical priorities. *Eur Heart J* 2012;**33**:176–182.
20. Rauch G, Jahn-Eimermacher A, Brannath W, Kieser M. Opportunities and challenges of combined effect measures based on prioritized outcomes. *Stat Med* 2014;**33**:1104–1120.
21. Rauch G, Rauch B, Schuler S, Kieser M. Opportunities and challenges of clinical trials in cardiology using composite primary endpoints. *World J Cardiol* 2015;**7**:1–5.
22. Subherwal S, Anstrom KJ, Jones WS, Felker MG, Misra S, Conte MS, Hiatt WR, Patel MR. Use of alternative methodologies for evaluation of composite end points in trials of therapies for critical limb ischemia. *Am Heart J* 2012;**164**:277–284.
23. Cleland JG. How to assess new treatments for the management of heart failure: composite scoring systems to assess the patients' clinical journey. *Eur J Heart Fail* 2002;**4**:243–247.
24. Dev S, Clare RM, Felker GM, Fiuzat M, Warner SL, O'Connor CM. Link between decisions regarding resuscitation and preferences for quality over length of life with heart failure. *Eur J Heart Fail* 2012;**14**:45–53.
25. Stevenson LW, Hellkamp AS, Leier CV, Sopko G, Koelling T, Warnica JW, Abraham WT, Kasper EK, Rogers JG, Califf RM, Schramm EE, O'Connor CM. Changing preferences for survival after hospitalization with advanced heart failure. *J Am Coll Cardiol* 2008;**52**:1702–1708.
26. Packer M. Proposal for a new clinical end point to evaluate the efficacy of drugs and devices in the treatment of chronic heart failure. *J Card Fail* 2001;**7**:176–182.
27. Redfield MM, Chen HH, Borlaug BA, Semigran MJ, Lee KL, Lewis G, LeWinter MM, Rouleau JL, Bull DA, Mann DL, Deswal A, Stevenson LW, Givertz MM, Ofili EO, O'Connor CM, Felker GM, Goldsmith SR, Bart BA, McNulty SE, Ibarra JC, Lin G, Oh JK, Patel MR, Kim RJ, Tracy RP, Velazquez EJ, Anstrom KJ, Hernandez AF, Mascette AM, Braunwald E. Effect of phosphodiesterase-5 inhibition on exercise capacity and clinical status in heart failure with preserved ejection fraction: a randomized clinical trial. *JAMA* 2013;**309**:1268–1277.
28. Abraham WT, Young JB, Leon AR, Adler S, Bank AJ, Hall SA, Lieberman R, Liem LB, O'Connell JB, Schroeder JS, Wheelan KR. 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**:2864–2868.
29. Packer M, Colucci W, Fisher L, Massie BM, Teerlink JR, Young J, Padley RJ, Thakkar R, Delgado-Herrera L, Salon J, Garratt C, Huang B, Sarapohja T. Effect of levosimendan on the short-term clinical course of patients with acutely decompensated heart failure. *JACC Heart Fail* 2013;**1**:103–111.

30. Hindricks G, Taborsky M, Glikson M, Heinrich U, Schumacher B, Katz A, Brachmann J, Lewalter T, Goette A, Block M, Kautzner J, Sack S, Husser D, Piorkowski C, Sogaard P. Implant-based multiparameter telemonitoring of patients with heart failure (IN-TIME): a randomised controlled trial. *Lancet* 2014;**384**:583–590.
31. Mozaffarian D, Benjamin EJ, Go AS, Arnett DK, Blaha MJ, Cushman M, de FS, Despres J, Fullerton HJ, Howard VJ, Huffman MD, Judd SE, Kissela BM, Lackland DT, Lichtman JH, Lisabeth LD, Liu S, Mackey RH, Matchar DB, McGuire DK, Mohler ER, III, Moy CS, Muntner P, Mussolino ME, Nasir K, Neumar RW, Nichol G, Palaniappan L, Pandey DK, Reeves MJ, Rodriguez CJ, Sorlie PD, Stein J, Towfighi A, Turan TN, Virani SS, Willey JZ, Woo D, Yeh RW, Turner MB. Heart Disease and Stroke Statistics—2015 update: a report from the American Heart Association. *Circulation* 2015;**131**:e29–e322.
32. Packer M, Narahara KA, Elkayam U, Sullivan JM, Pearle DL, Massie BM, Creager MA. Double-blind, placebo-controlled study of the efficacy of flosequinan in patients with chronic heart failure. Principal Investigators of the REFLECT Study. *J Am Coll Cardiol* 1993;**22**:65–72.
33. Galie N, Hoeper MM, Humbert M, Torbicki A, Vachiery JL, Barbera JA, Beghetti M, Corris P, Gaine S, Gibbs JS, Gomez-Sanchez MA, Jondeau G, Klepetko W, Opitz C, Peacock A, Rubin L, Zellweger M, Simonneau G. Guidelines for the diagnosis and treatment of pulmonary hypertension: the Task Force for the Diagnosis and Treatment of Pulmonary Hypertension of the European Society of Cardiology (ESC) and the European Respiratory Society (ERS), endorsed by the International Society of Heart and Lung Transplantation (ISHLT). *Eur Heart J* 2009;**30**:2493–2537.
34. Taichman DB, Ornelas J, Chung L, Klinger JR, Lewis S, Mandel J, Palevsky HI, Rich S, Sood N, Rosenzweig EB, Trow TK, Yung R, Elliott CG, Badesch DB. Pharmacologic therapy for pulmonary arterial hypertension in adults: CHEST guideline and expert panel report. *Chest* 2014;**146**:449–475.
35. Anker SD, Agewall S, Borggrefe M, Calvert M, Jaime CJ, Cowie MR, Ford I, Paty JA, Riley JP, Swedberg K, Tavazzi L, Wiklund I, Kirchhof P. The importance of patient-reported outcomes: a call for their comprehensive integration in cardiovascular clinical trials. *Eur Heart J* 2014;**35**:2001–2009.
36. Food and Drug Administration. *Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims. Guidance for Industry* 2009.
37. European Medicines Agency. Innovative drug development approaches: final report from the EMEA/CHMP think-tank group on innovative drug development. http://www.ema.europa.eu/docs/en_GB/document_library/Other/2009/10/WC500004913.pdf (30 April 2015).
38. European Medicines Agency. Pilot project on adaptive licensing. http://www.ema.europa.eu/docs/en_GB/document_library/Other/2014/03/WC500163409.pdf (26 June 2014).
39. Eichler HG, Oye K, Baird LG, Abadie E, Brown J, Drum CL, Ferguson J, Garner S, Honig P, Hukkelhoven M, Lim JC, Lim R, Lumpkin MM, Neil G, O'Rourke B, Pezalla E, Shoda D, Seyfert-Margolis V, Sigal EV, Sobotka J, Tan D, Unger TF, Hirsch G. Adaptive licensing: taking the next step in the evolution of drug approval. *Clin Pharmacol Ther* 2012;**91**:426–437.
40. European Medicines Agency. Guideline on good pharmacovigilance practices (GVP): Module VIII Post authorisation safety studies (Rev 1). http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2012/06/WC500129137.pdf (30 March 2015).
41. Rosano GMC, Anker S, Marrocco W, Coats AJS. Adaptive licensing: a way forward in the approval process of new therapeutic agents in Europe. *Clin Trials Regul Sci Cardiol* 2015;**1**:1–2.
42. Anker SD, McMurray JJ. Time to move on from 'time-to-first': should all events be included in the analysis of clinical trials? *Eur Heart J* 2012;**33**:2764–2765.
43. Dunlay SM, Shah ND, Shi Q, Morlan B, VanHouten H, Long KH, Roger VL. Lifetime costs of medical care after heart failure diagnosis. *Circ Cardiovasc Qual Outcomes* 2011;**4**:68–75.
44. Bradley EH, Curry LA, Krumholz HM. Strategies to reduce heart failure readmissions. *JAMA* 2014;**311**:1160.
45. Joynt KE, Jha AK. A path forward on Medicare readmissions. *N Engl J Med* 2013;**368**:1175–1177.
46. Rogers JK, McMurray JJ, Pocock SJ, Zannad F, Krum H, van Veldhuisen DJ, Swedberg K, Shi H, Vincent J, Pitt B. Eplerenone in patients with systolic heart failure and mild symptoms: analysis of repeat hospitalizations. *Circulation* 2012;**126**:2317–2323.
47. Borer JS, Bohm M, Ford I, Komajda M, Tavazzi L, Sendon JL, Alings M, Lopez-de-Sa E, Swedberg K. Effect of ivabradine on recurrent hospitalization for worsening heart failure in patients with chronic systolic heart failure: the SHIFT Study. *Eur Heart J* 2012;**33**:2813–2820.
48. Rogers JK, Pocock SJ, McMurray JJ, Granger CB, Michelson EL, Ostergren J, Pfeffer MA, Solomon SD, Swedberg K, Yusuf S. Analysing recurrent hospitalizations in heart failure: a review of statistical methodology, with application to CHARM-Preserved. *Eur J Heart Fail* 2014;**16**:33–40.
49. Rogers JK, Jhund PS, Perez AC, Bohm M, Cleland JG, Gullestad L, Kjekshus J, van Veldhuisen DJ, Wikstrand J, Wedel H, McMurray JJ, Pocock SJ. Effect of rosuvastatin on repeat heart failure hospitalizations: the CORONA Trial (Controlled Rosuvastatin Multinational Trial in Heart Failure). *JACC Heart Fail* 2014;**2**:289–297.
50. Amorim LD, Cai J. Modelling recurrent events: a tutorial for analysis in epidemiology. *Int J Epidemiol* 2015;**44**:324–333.
51. Abraham WT, Adamson PB, Bourge RC, Aaron MF, Costanzo MR, Stevenson LW, Strickland W, Neelagaru S, Raval N, Krueger S, Weiner S, Shavelle D, Jeffries B, Yadav JS. Wireless pulmonary artery haemodynamic monitoring in chronic heart failure: a randomised controlled trial. *Lancet* 2011;**377**:658–666.
52. Rogers, J. Recurrent events analysis, not so straightforward! http://www.statisticsviews.com/details/feature/5786631/Recurrent-events-analysis-not-so-straightforward.html (28 January 2014).
53. Novartis Pharmaceuticals. Efficacy and safety of LCZ696 compared to valsartan on morbidity and mortality in heart failure patients with preserved ejection fraction (PARAGON-HF). https://clinicaltrials.gov/ct2/show/NCT01920711 (30 May 2015).
54. Committee for Medicinal Products for Human Use (CHMP). Guideline on the choice of the non-inferiority margin. EMEA/CPMP/EWP/21 58/99 2005. http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/09/WC500003417.pdf
55. European Medicines Agency Committee for Medicinal Products for Human Use (CHMP). Guideline on missing data in confirmatory clinical trials (EMA/CPMP/EWP/1776/99). http://www ema europa eu/docs/en_GB/document_library/Scientific_guideline/2010/09/WC500096793 pdf 2010.
56. Senn S. Disappointing dichotomies. *Pharmaceut Stat* 2003;**2**:239–240.
57. Lewis JA. In defence of the dichotomy. *Pharmaceut Stat* 2004;**3**:77–79.
58. Hawkins NM, Petrie MC, Burgess MI, McMurray JJ. Selecting patients for cardiac resynchronization therapy: the fallacy of echocardiographic dyssynchrony. *J Am Coll Cardiol* 2009;**53**:1944–1959.
59. Snapinn SM, Jiang Q. Responder analyses and the assessment of a clinically relevant treatment effect. *Trials* 2007;**8**:31.
60. Vaduganathan M, Greene SJ, Ambrosy AP, Gheorghiade M, Butler J. The disconnect between phase II and phase III trials of drugs for heart failure. *Nat Rev Cardiol* 2013;**10**:85–97.
61. Hayward RA, Kent DM, Vijan S, Hofer TP. Multivariable risk prediction can greatly enhance the statistical power of clinical trial subgroup analysis. *BMC Med Res Methodol* 2006;**6**:18.
62. Kent DM, Rothwell PM, Ioannidis JP, Altman DG, Hayward RA. Assessing and reporting heterogeneity in treatment effects in clinical trials: a proposal. *Trials* 2010;**11**:85.