
There may be differences between this version and the published version. You are advised to consult the publisher’s version if you wish to cite from it.

http://eprints.gla.ac.uk/120856/

Deposited on: 20 September 2017
Traditional and New Composite Endpoints in Heart Failure Clinical Trials: Facilitating Comprehensive Efficacy Assessments and Improving Trial Efficiency

Stefan D. Anker1; Stefan Schroeder2; Dan Atar3; Jeroen J. Bax4; Claudio Ceconi5; Martin R. Cowie6; Adam Crisp7; Fabienne Dominjon8; Ian Ford9; Hossein-Ardeschir Ghofrani10-12; Savion Gropper13; Gerhard Hindricks14; Mark A. Hlatky15; Richard Holcomb16; Narimon Honarpour17; J. Wouter Jukema4; Albert M. Kim18; Michael Kunz2; Martin Lefkowitz19; Chantal Le Floc’h8; Ulf Landmesser20; Theresa A. McDonagh21; John J. McMurray22; Bela Merkely23; Milton Packer24; Krishna Prasad25; James Revkin18; Giuseppe Rosano26, 27; Ransi Somaratne17; Wendy Gattis Stough28; Adriaan A. Voors29; Frank Ruschitzka30

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

Corresponding Author: Stefan D. Anker, MD PhD; Division of Innovative Clinical Trials, Department of Cardiology & Pneumology, University Medical Center Göttingen (UMG); Robert-Koch-Str. 40; D-37075 Göttingen, Germany; Email: s.anker@cachexia.de
Abstract

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 enroll 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
INTRODUCTION

Composite endpoints are increasingly used as primary efficacy measures in heart failure clinical trials (Supplemental Appendix Table 1)\(^1\) 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 ten 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 analytic 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 two-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.\textsuperscript{9,10} Typically, reporting the individual components and testing the treatment effect on each is sufficient.

Similar logic applies to all-cause versus 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 program.\textsuperscript{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, CI 0.79-0.97, p=0.012).\textsuperscript{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 versus 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 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 analyzed independently because of the problem of competing risks. Composite endpoints can solve this dilemma if they include both fatal and non-fatal events. However, analyzing 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,\textsuperscript{13} which decreases the relevance (not confidence) of any finding. For example, in the Systolic Heart Failure Treatment with the \textit{I} \textit{f} Inhibitor Ivabradine Trial (SHIFT) study, the primary endpoint was a composite of cardiovascular death or hospital admission for worsening heart failure.\textsuperscript{14} Ivabradine reduced the risk of the primary endpoint compared with placebo in patients with a heart rate $\geq 70$ beats per minute (HR 0.82, CI 0.75-0.90, $p<0.0001$), mainly driven by a reduction in heart failure hospitalizations (16\% vs. 21\%, HR 0.74, CI 0.66-0.83, $p<0.0001$) with no effect on cardiovascular mortality (14\% vs. 15\%, HR 0.91, CI 0.80-1.03, $p=0.13$).\textsuperscript{14} The European Medicines Agency (EMA) approved ivabradine, but only for patients with heart rates above 75 beats per minute where a possible nominal reduction in overall mortality was observed.\textsuperscript{15} Food and Drug Administration (FDA) approval was recently granted for a reduction in heart failure hospitalization in patients with a heart rate $\geq 70$ beats per minute. 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: 1) all-cause mortality (relative risk 1.02, CI 0.88-1.18, $p=0.80$) and 2) the composite of all-cause mortality or cardiac arrest with resuscitation, hospitalization for heart failure, or administration of intravenous inotropes or vasodilators for $\geq 4$ hours without hospitalization (relative risk 0.87, 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.\textsuperscript{16} Statistical methods to weight outcomes according to severity have been proposed in heart failure and other disease states.\textsuperscript{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. A treatment may be worthwhile and considered valuable by patients when it improves their clinical status even if it does not prolong their survival. A clinical status composite endpoint has been developed and used 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. 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.$^{24,25}$ 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 likely 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-minute walk test distance has been the primary endpoint in most registration trials of pulmonary artery hypertension,\(^{33;34}\) accepted by both 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 Minnesota Living With Heart Failure Score (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 labeling claims.\(^{35;36}\) While health-related quality of life assessments are clinically relevant and can be informative, the methodologic 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.\textsuperscript{35} Whether regulatory agencies would accept the MLWHF or KCCQ as supportive evidence of efficacy remains to be seen and would likely 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-minute 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.

\textbf{Role of Adaptive Licensing}

In 2007, EMA described its openness to innovative drug development approaches.\textsuperscript{37} One such initiative is the Pilot Project on Adaptive Licensing.\textsuperscript{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”.\textsuperscript{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.\textsuperscript{38-40} This process fulfills 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 license. 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. 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, 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 ANALYTIC 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.\textsuperscript{42} Hospitalization is the major contributor to the overall cost of heart failure care,\textsuperscript{43} which has led to targeted interventions to reduce readmissions.\textsuperscript{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.\textsuperscript{46-49} Several approaches to recurrent event analyses have been tested using major heart failure trial datasets.\textsuperscript{46-51} All have limitations, as complex assumptions are made in determining the study sample size and in the statistical modelling.\textsuperscript{52} One concern is overestimation of the treatment effect,\textsuperscript{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,\textsuperscript{48;50} but none of the methods have 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.\textsuperscript{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.\textsuperscript{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.\textsuperscript{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, left ventricular size, ejection fraction, 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.\textsuperscript{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 favorable phase II results did not translate into improved outcome in phase III.\textsuperscript{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 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.\textsuperscript{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 over interpretation 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 on patients is broader than hospitalizations and death, and it will be important to conduct trials that reflect endpoints important to patients\textsuperscript{35} (and payers). Further, the declining resources for conducting larger, longer, and more costly trials is a reality that cannot be ignored.\textsuperscript{5} Composites that reflect both clinical status and traditional “hard” events, new analytic 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.

Conflicts of Interest

Stefan D. Anker: Research grants from Vifor and Abbott Vascular; personal fees from Vifor, Abbott Vascular, Bayer Pharma, Novartis, Lonestar Heart, Respircardia, ZS Pharma, Relypsa, Biotronik, Cardiorentis, Servier

Stefan Schroeder: Employee of Bayer Healthcare

Dan Atar: Personal fees from Astra-Zeneca, Bayer Healthcare, Boehringer-Ingelheim, BMS/Pfizer, Novartis, Vifor

Jeroen J. Bax: Nothing to disclose

Claudio Ceconi: Nothing to disclose

Martin R. Cowie: Research grants from ResMed, Boston Scientific, and Bayer; personal fees from ResMed, Boston Scientific, Bayer, Servier, Novartis, St. Jude Medical, and Pfizer
Adam Crisp: Employee of GlaxoSmithKline

Fabienne Dominjon: Employee of Servier

Ian Ford: Nothing to disclose

Hossein-Ardeschir Ghofrani: Personal fees (board membership, consultancy, paid lectures, or travel expenses) from Actelion, Bayer, GSK, Novartis, Pfizer, Bellerophon Therapeutics, United Therapeutics; research grants from Actelion, Bayer, Pfizer, Novartis.

Savion Gropper: Employee of Boehringer-Ingelheim GmBH & Co.KG

Gerhard Hindricks: Research grants from St. Jude Medical, Biotronik, Boston Scientific

Mark A. Hlatky: Nothing to disclose

Richard Holcomb: Nothing to disclose

Narimon Honarpour: Employee of Amgen

J. Wouter Jukema: Research grants from and CME speaker 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, European Community Framework KP7 Programme

Albert M. Kim: Employee of Pfizer

Michael Kunz: Employee of Bayer Healthcare

Martin Lefkowitz: Employee of Novartis Pharmaceuticals Company

Chantal Le Floch: Employee of Servier

Ulf Landmesser: Research grants from Pfizer and Roche; honoraria (lecture and advisory) from Roche, MSD, Sanofi, and Amgen

Theresa A. McDonagh: Personal fees (honoraria) from Novartis, Vifor, ZS Pharma
John J. McMurray: Nothing to disclose

Bela Merkely: Personal fees (lectures and advisory board membership) from Orion Pharma

Milton Packer: Nothing to disclose

Krishna Prasad: Nothing to disclose.

James Revkin: Nothing to disclose.

Giuseppe Rosano: Nothing to disclose

Ransi Somaratne: Employee and shareholder of Amgen, Inc.

Wendy Gattis Stough: Personal fees from European Society of Cardiology, Relypsa, Covis, Stealth Peptides

Adriaan A. Voors: Research grant 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; non-financial support from Alere and Merck/MSD

Frank Ruschitzka: Nothing to disclose
References


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.


52. Rogers, J. Recurrent events analysis, not so straightforward!


57. Lewis JA. In defence of the dichotomy. Pharmaceut Statist 2004;3:77-79.


Table 1. Advantages and Disadvantages of Composite Endpoints

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

HF = heart failure; MI = myocardial infarction
Table 2. Characteristics of Robust Composite Endpoints

<table>
<thead>
<tr>
<th>A Composite Endpoint should:</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Provide reliable and precise estimates of efficacy and safety</td>
</tr>
<tr>
<td>2. Be clinically meaningful or relevant to physicians, patients, and care providers in terms of</td>
</tr>
<tr>
<td>characterising disease progression, stabilization, or reversal</td>
</tr>
<tr>
<td>3. Meaningfully characterise the burden of disease for patients</td>
</tr>
<tr>
<td>4. Yield information that could be used in conjunction with other data to determine societal</td>
</tr>
<tr>
<td>valuation of new therapies or interventions</td>
</tr>
<tr>
<td>5. Improve the efficiency of clinical trials while maintaining high validity and quality</td>
</tr>
<tr>
<td>Topic</td>
</tr>
<tr>
<td>-----------------------------</td>
</tr>
</tbody>
</table>
| 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 small proportion of patients).  
2. How to confirm accuracy of assumptions and simulations used in modeling?  
3. Involve regulators in planning modeling 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. |
Supplemental Appendix Table 1. Composite Endpoints in Heart Failure Clinical Trials

<table>
<thead>
<tr>
<th>Trial</th>
<th>Primary Endpoint</th>
</tr>
</thead>
<tbody>
<tr>
<td>CONSENSUS</td>
<td>All-cause mortality</td>
</tr>
<tr>
<td>SOLVD-P</td>
<td>All-cause mortality</td>
</tr>
<tr>
<td>SOLVD-T</td>
<td>All-cause mortality</td>
</tr>
<tr>
<td>V-HeFT II</td>
<td>All-cause mortality</td>
</tr>
<tr>
<td>Val-HeFT</td>
<td>Co-primary endpoints</td>
</tr>
<tr>
<td></td>
<td>All-cause mortality</td>
</tr>
<tr>
<td></td>
<td>Mortality and morbidity (cardiac arrest with resuscitation, hospitalization for</td>
</tr>
<tr>
<td></td>
<td>heart failure, administration of IV inotropic or vasodilator drugs for ≥4 hours</td>
</tr>
<tr>
<td></td>
<td>without hospitalization)</td>
</tr>
<tr>
<td>CHARM</td>
<td>Overall program: All-cause mortality</td>
</tr>
<tr>
<td></td>
<td>Individual trials: Cardiovascular death or heart failure hospitalization</td>
</tr>
<tr>
<td>ELITE II</td>
<td>All-cause mortality</td>
</tr>
<tr>
<td>MERIT-HF</td>
<td>All-cause mortality</td>
</tr>
<tr>
<td>COPERNICUS</td>
<td>All-cause mortality</td>
</tr>
<tr>
<td>CIBIS-II</td>
<td>All-cause mortality</td>
</tr>
<tr>
<td>BEST</td>
<td>All-cause mortality</td>
</tr>
<tr>
<td>A-HeFT</td>
<td>Composite score made up of weighted values for death from any cause, a first</td>
</tr>
<tr>
<td></td>
<td>hospitalization for heart failure during the 18-month follow-up period, and change</td>
</tr>
<tr>
<td></td>
<td>in the quality of life at 6 months.</td>
</tr>
</tbody>
</table>
Supplemental Appendix Table 1. CompositeEndpoints in Heart Failure Clinical Trials

<table>
<thead>
<tr>
<th>Trial</th>
<th>Primary Endpoint</th>
</tr>
</thead>
<tbody>
<tr>
<td>PRAISE&lt;sup&gt;74&lt;/sup&gt;</td>
<td>All-cause mortality or cardiovascular morbidity (defined as hospitalization for at least 24 hours for acute pulmonary edema, severe hypoperfusion, acute myocardial infarction, or sustained or hemodynamically destabilizing ventricular tachycardia or fibrillation)</td>
</tr>
<tr>
<td>PRAISE II&lt;sup&gt;75&lt;/sup&gt;</td>
<td>All-cause mortality</td>
</tr>
<tr>
<td>RALES&lt;sup&gt;76&lt;/sup&gt;</td>
<td>All-cause mortality</td>
</tr>
<tr>
<td>SHIFT&lt;sup&gt;14&lt;/sup&gt;</td>
<td>Cardiovascular death or HF hospitalization</td>
</tr>
<tr>
<td>EMPHASIS-HF&lt;sup&gt;77&lt;/sup&gt;</td>
<td>Cardiovascular death or HF hospitalization</td>
</tr>
<tr>
<td>TOPCAT&lt;sup&gt;78&lt;/sup&gt;</td>
<td>Composite of death from CV causes, aborted cardiac arrest, or hospitalization for the management of heart failure</td>
</tr>
<tr>
<td>PARADIGM-HF&lt;sup&gt;79&lt;/sup&gt;</td>
<td>Cardiovascular death or HF hospitalization</td>
</tr>
<tr>
<td>MADIT-CRT&lt;sup&gt;80&lt;/sup&gt;</td>
<td>All-cause mortality or non-fatal HF events (signs/symptoms consistent with HF that was responsive to IV decongestive therapy on an outpatient basis or an augmented decongestive regimen with oral or parenteral medications during an in-hospital stay)</td>
</tr>
<tr>
<td>RAFT&lt;sup&gt;81&lt;/sup&gt;</td>
<td>All-cause mortality or HF leading to hospitalization</td>
</tr>
<tr>
<td>SCD-HeFT&lt;sup&gt;82&lt;/sup&gt;</td>
<td>All-cause mortality</td>
</tr>
<tr>
<td>OPTIME&lt;sup&gt;83&lt;/sup&gt;</td>
<td>Cumulative days of hospitalization for cardiovascular causes</td>
</tr>
<tr>
<td>EVEREST Outcome&lt;sup&gt;84&lt;/sup&gt;</td>
<td>Dual primary endpoints of all-cause mortality and</td>
</tr>
</tbody>
</table>
Supplemental Appendix Table 1. Composite Endpoints in Heart Failure Clinical Trials

(continued)

<table>
<thead>
<tr>
<th>Trial</th>
<th>Primary Endpoint</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>cardiovascular death or hospitalization for heart failure</td>
</tr>
<tr>
<td>ASCEND&lt;sup&gt;85&lt;/sup&gt;</td>
<td>Two co-primary endpoints: change in self-reported dyspnea at 6 and 24 hours after study-drug initiation and the composite endpoint of rehospitalization for HF and death from any cause</td>
</tr>
</tbody>
</table>
| VERITAS<sup>86</sup> | Change from baseline in dyspnea over first 24 hours of treatment (individual studies)  
Incidence of death or worsening HF at 7 days (combined studies) |
| REVIVE<sup>29</sup> | Composite clinical score: Improved (patients considered themselves moderately or markedly improved at all pre-specified time points and showed no evidence of clinical deterioration); Unchanged (not improved or worse); Worse (died within 5 days, persistent or unresponsive HF symptoms after first 24 hrs of randomized therapy or worsening HF anytime within 5 days that required a rescue intervention specifically to relieve the symptoms, or patients considered themselves to have moderately or markedly worsened on global assessment at any of the follow-up time points). |
| SURVIVE<sup>87</sup> | All-cause mortality at 180 days                                                   |
| RELAX-AHF<sup>88</sup> | Two primary efficacy endpoints: change in patient-reported dyspnea by AUC of the VAS scores from baseline to day 5; |
### Supplemental Appendix Table 1. Composite Endpoints in Heart Failure Clinical Trials

(continued)

<table>
<thead>
<tr>
<th>Trial</th>
<th>Primary Endpoint</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>moderately or markedly improved patient reported dyspnea relative to the start of study drug assessed by 7-level Likert scale at 6, 12, and 24 hours</td>
</tr>
<tr>
<td>RELAX-AHF-2\textsuperscript{39}</td>
<td>Time to confirmed cardiovascular death through 180 days and time to worsening heart failure through 5 days (also considering death), with all-cause mortality through 180 days pre-specified as a safety endpoint.</td>
</tr>
<tr>
<td>TRUE-HF\textsuperscript{30}</td>
<td>Two primary efficacy endpoints: 1) improvement in a hierarchical clinical composite endpoint (assessed at 6, 24, and 48 hours after start of study drug) including patient global assessment of symptomatic improvement, lack of improvement, or worsening; persistent or worsening heart failure requiring pre-specified mechanical or pharmacologic interventions, and all-cause death; 2) freedom from cardiovascular death throughout the duration of the trial.</td>
</tr>
</tbody>
</table>

AUC = area under the curve; CV = cardiovascular; HF = heart failure; VAS = visual analog scale