Published in final edited form as:

Heart Fail Clin. 2013 January ; 9(1): 15–28. doi:10.1016/j.hfc.2012.09.002.

Endpoints for Comparative Effectiveness Research in Heart Failure

Larry A. Allen, MD, MHS* and John A. Spertus, MD, MPH†

*Colorado Cardiovascular Outcomes Research Consortium, Division of Cardiology, Department of Medicine, University of Colorado Denver, Aurora, Colorado

†Saint Luke's Mid America Heart Institute and the University of Missouri – Kansas City, Kansas City, MO

Abstract

With the increasing availability of therapeutic strategies (drugs, devices, disease management systems) and the growing complexity of health care delivery, there is an attendant need for objective evidence of the tangible benefits of different approaches to care. This is particularly true for patients with heart failure, a common, morbid, and resource-intensive disease. There are few well-proven therapies for patients with acute decompensation or for patients with normal LVEF. Comparative effectiveness research (CER) offers an important avenue for making progress in the field. However, CER, like any well-designed research program, requires the explicit articulation of clinically important outcomes to be compared. For patients with heart failure, there is a need to develop endpoint measures that capture the totality of potential benefits and risks for alternative therapeutic approaches. Ultimately, for one therapeutic approach to be considered superior to another, it must improve one of three relevant endpoints: make patients live longer, make them feel better, or save money without adversely affecting the other two goals. Importantly, these outcomes must be measured directly and surrogates should be avoided, even if such surrogates appear to be associated with clinically meaningful, patient-centered outcomes. In this review, we discuss the available CER endpoint domains from both a clinical and a statistical perspective, summarize the wide variety of endpoints used in CER studies, and suggest steps for greater standardization of endpoints across CER studies of patients with heart failure.

Keywords

Endpoints/end points; congestive heart failure; comparative effectiveness research; study design

INTRODUCTION

As medical therapies continue to expand and the delivery of health care becomes increasingly complex, there is a growing need to inform patients, clinicians, and policy

© 2012 Elsevier Inc. All rights reserved.

Corresponding Author: Larry A. Allen, MD, MHS, Academic Office 1, 12631 East 17th Avenue, Mailstop B130, PO Box 6511, Aurora, Colorado 80045. larry.allen@ucdenver.edu. Phone: 303-724-4713. Fax: 303-724-2094.

Disclosures: Dr. John Spertus owns the copyright to the Kansas City Cardiomyopathy Questionnaire. There are no other relevant conflicts to report

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

makers of the relative value of alternative treatment options. Comparative effectiveness research (CER) is emerging as a central approach to help guide health care decisions by providing data about the relative safety and effectiveness of various therapeutic approaches. While CER offers the hope of improving the quality of our health care delivery system, it is a science in rapid evolution.

Ultimately, the process of comparing one therapy to another involves comparing how various approaches differentially affect clinically meaningful outcomes. From the patient, clinician, and societal perspectives, the value of a therapy stems from its impact on three main concepts: Does the therapy make patients live longer? Does the therapy make patients feel better? Or does the therapy save resources? Although survival, quality of life, and cost may appear on face value to be relatively straightforward concepts, accurate assessment and interpretation of these endpoints is complex. Researchers must choose which measures to collect, how to collect those measures efficiently and accurately, and how to weigh competing measures against one another.

Heart failure exemplifies the issues around endpoints in CER.¹⁻³ The heart failure syndrome includes patients with diverse presentations and pathophysiology, usually in combination with high degrees of comorbidity and advanced age. Heart failure is the leading cause of hospitalization and rehospitalization among older adults in the United States, at a direct cost of >\$40 billion per year.⁴ Heart failure remains a common, morbid, and resource intensive disease, in part because very few treatment approaches have been shown to be efficacious for patients with acute decompensation⁵ or for patients with heart failure and normal left ventricular ejection fraction (LVEF).⁶ Additionally, because heart failure is a highly symptomatic disease, and because many attempted therapies have failed to improve mortality, there is increasing interest in assessing effectiveness for a wide range of endpoints that reflect quality of life and cost. As such, the field of heart failure research is marked by significant heterogeneity in endpoint assessment.

CER for patients with heart failure needs to develop endpoints that encompass the totality of potential benefits (and risks) for alternative therapeutic approaches, including relieving symptoms, improving overall quality of life, limiting hospitalization (a surrogate for costs), and prolonging survival. Greater standardization of endpoints across CER studies would allow therapies to be more easily compared and prioritized by patients, clinicians, and payers. In this review, we discuss the potential CER endpoint domains from both a clinical and a research perspective, summarize the wide variety of endpoints used in completed and ongoing CER studies, and suggest steps for greater standardization of endpoints across CER for patients with heart failure.

DEFINITIONS AND OVERALL CONSIDERATIONS

Endpoints in medical research are specific, identifiable, downstream events or changes in a patient's condition. Clinically meaningful endpoints should be evident to and important to patients (e.g. not laboratory measures). Additionally, ideal endpoints should be unambiguous, consistent, easy to obtain, inexpensive to measure, sensitive to the processes of care that they are designed to reflect, and simple to understand. Heart failure is typically progressive, systemic in its effects, highly symptomatic, and lethal. Therefore, a wide range of endpoint options are available (Table 1). When considering endpoint design for heart failure studies a general framework of the disease process should be taken into account.

While there are diverse and complex pathophysiological disturbances in heart failure, including left ventricular remodeling and neurohormonal dysregulation, the outcomes of mortality and health status (symptoms, function, and quality of life) are relatively constant across the spectrum of the heart failure syndrome. In fact, the definition of heart failure as a

syndrome, rather than a disease, underscores the concept that multiple pathophysiological disturbances lead to similar health status limitations and mortality rates. Regardless of its etiology, heart failure patients follow a similar, albeit variable, clinical course. Figure 1 outlines a schematic approach to the syndrome. Patients are often diagnosed with heart failure after suffering with undiagnosed symptoms for some time (stage 1). After initial diagnosis and treatment, they often enter a period of relative stability (stage 2) punctuated with periodic but increasing episodes of decompensation (stage 3) resulting in acute care/hospitalizations. At some point, they enter a terminal trajectory (stage 4) at which heart transplantation of left ventricular assist devices can be considered for a very select group of patients with reduced LVEF to move them to a previous phase of their trajectory (e.g. stage 2). Ultimately, however, a terminal stage emerges (stage 5) for which the therapeutic goals shift towards comfort care in preparation for death. This progressive course may be truncated at any point by sudden cardiac death, which is relatively common among patients with heart failure and reduced LVEF. This schematic, originally created by Dale Renlund and colleagues, highlights the variable, but generally progressive course of the disease.⁷ It also highlights clinically important outcomes—health status, hospitalizations, transplant or device therapy, and death—that can be measured throughout the disease course.

VITAL STATUS

Vital status—variably expressed as mortality or survival—is of unquestioned importance as a clinically relevant endpoint. Patients with symptomatic heart failure have an average survival of less than 5 years,⁸ so vital status measures are particularly important in assessing treatments for these patients. Unlike many other endpoints, all-cause mortality is objective, relatively easy to assess, and can often be obtained retrospectively. Consequently, mortality is generally considered the “gold standard” of clinical endpoints, and it is the most commonly reported endpoint in CER studies.

Vital status can be obtained through a variety of mechanisms, including queries to families, to health plans, or to administrative records (State death certificates, the National Death Index, and Social Security Death Master File), if adequate patient identifiers (e.g. social security numbers) are available.

Controversy remains, however, about whether it is preferable to measure all-cause or disease-specific mortality.⁹ All-cause mortality has the advantage of being completely objective and has a higher overall event rate than disease-specific mortality. The critique of all-cause mortality, however, is that it will include events that are unlikely to be responsive to the heart failure treatments in question, so the ‘signal’ of treatment efficacy may be obscured by the ‘noise’ of other etiologies for death. Therefore, disease-specific event rates related to the specific mechanism of action of the therapy (e.g., sudden death in defibrillator studies) are more likely to demonstrate a treatment effect, assuming that the mechanism of action of the therapy is reasonably well understood. Mode of death may also significantly influence quality of life and cost in the last months of life (e.g. sudden cardiac death as compared with progressive pump failure). It should be noted that therapies that reduce one mode of death may inadvertently increase another (e.g., arrhythmic death v. non-arrhythmic death with the implantation of defibrillators immediately after myocardial infarction),¹⁰ and thus even if a disease-specific cause of death is chosen, all-cause mortality should still be reported. Data from heart failure trials among patients with reduced LVEF have found that most post-discharge deaths are cardiovascular in nature,^{11, 12} even among older patients and those with normal LVEF.^{13, 14} If disease-specific mortality is felt to be an important endpoint, adjudication of cause-specific mortality is needed.¹⁵ Death certificates are notoriously inaccurate.¹⁶ Therefore, determining deaths due to cardiovascular causes, heart failure, progressive pump failure, sudden cardiac death, etc. invariably require some active

interpretation of the clinical events leading up to death. This necessitates prospective standardized capture of physician-determined cause of death or detailed chart reviews to adjudicate cause of death. When patients die outside of the hospital, arbitrary rules must be established to assign the cause of their demise; often clearly non-cardiovascular causes of death are labeled as such (e.g. trauma or metastatic cancer) and then all other deaths are presumed to be cardiovascular in nature. All of these factors should be considered when deciding whether or not to measure disease-specific mortality. In the end, observational CER studies, especially those using administrative data, lack detailed clinical information, so all-cause mortality is the preferred endpoint.

Despite its obvious importance, death as a single endpoint has the disadvantage of limited statistical power and misses other dimensions of outcome important to patients. Although heart failure is a highly lethal condition, use of mortality as the endpoint still requires large sample sizes, longer follow-up, and relatively large treatment effect, to show significant differences. The heterogeneity of the population, the high prevalence of contributing comorbidities, and the failure of many therapies to improve survival has discouraged the use of mortality as the lone endpoint in clinical trials. Although some administrative heart failure databases, which include tens of thousands of patients, can overcome issues of statistical power that plague clinical trials, they cannot address many of the other methodological concerns (e.g. treatment selection bias, confounding, etc.) present in observational studies of treatment effect.

HEALTH STATUS

A major limitation in the current focus on survival is that it incorrectly presumes that all survival represents an equally favorable outcome. Health status—symptoms, function, and quality of life—are primary concerns for patients. In fact, multiple studies have demonstrated that nearly half of patients with heart failure would make healthcare decisions that improve their quality of life even if it reduced their life expectancy.^{17–19} Near the end of life, health status may be given even greater priority. While traditionally considered a ‘soft’ outcome, since it comes from patients rather than an ‘objective’ test, the reproducibility and sensitivity to clinical change of standardized questionnaires is better than many other modalities of quantifying patients’ function.²⁰ Yet, despite the availability of validated measures, a principal challenge in conducting CER with health status as an outcome has been the lack of systematic quantification of health status in clinical studies, observational research, and clinical care.

Survival and quality of life both represent critical patient-centered outcomes, but their practical use as endpoints in research is quite disparate. Whereas all-cause mortality is a simple dichotomous measure that can usually be obtained retrospectively, quality of life measures must be prospectively obtained via direct input from the patient. Thus, while quality of life is clearly an important endpoint, its use in CER studies has been relatively limited due to these logistical limitations. Because health care value is receiving increasing attention and because value cannot be determined without including some measure of quality of life (e.g. resource use per quality adjusted life years [cost/QALY] for cost-utility analyses), CER studies are likely to increasingly involve assessments of quality of life.

Methods for Assessing Health Status

Heart failure impacts patients’ lives in a number of ways. While many use the term ‘quality of life’ to encompass all outcomes beyond survival, we consider it to be a distinct domain and prefer the term ‘health status’ to refer to the combination of domains in which a disease can impact patients. These include the following: the symptoms of disease; the physical, emotional and social limitations associated with these symptoms and patients’ quality of

life; and the discrepancy between patients' current symptoms and functioning and how they expect to be feeling and functioning. Figure 2 outlines a framework for understanding how heart failure can impact the health status of afflicted patients.

In clinical research studies, there are several methods for quantifying the impact of the disease on patients' health status. Prototypical examples of health status measures are provided in Table 2. An important consideration is whether generic health status measures, that quantify a patients' overall health, or disease-specific measures, that focuses upon the clinical syndrome of heart failure, should be used.

Disease-Specific Health Status Measures

While there are a myriad of potential etiologies for heart failure, patients are unaware of these pathophysiological mechanisms. What they are aware of are the symptoms of fatigue, shortness of breath, and edema. These symptoms, in turn, can limit their physical and social activities and impair their emotional functioning. As noted above, the degree to which their symptoms and functional limitations differ from their expectations for their health, the worse their quality of life will be.

The most common metric of the impact of heart failure on patients is the NYHA functional classification.²¹ This 4-class measure is the most commonly used measure of patients' symptoms and function in clinical practice and in research, but suffers from several critical limitations. First, it reflects patients' health status from physicians' perspectives and not patients themselves. Second, its validity and reproducibility have been questioned.^{22, 23} Less commonly used metrics, particularly in trials of acute heart failure, are single-item measures of dyspnea or global health.²⁴ Currently, there are no validated instruments for dyspnea assessment that are accurate, reliable, reproducible between observers, and sensitive to acute changes in dyspnea.²⁵ Instead, an assortment of poorly validated instruments are used assessing dyspnea, including Likert scales, visual analog scales, and more complicated measures.^{26, 27} Patterns of dyspnea resolution are significantly affected by choice of response instrument.²⁸

A number of disease-specific health status measures have been developed for patients with heart failure, of which the most commonly used are the Minnesota Living with Heart Failure Questionnaire (MLWHF)²⁹ and the Kansas City Cardiomyopathy Questionnaire (KCCQ).³⁰ The MLWHF provides an overall summary score, with a physical and mental component. The KCCQ follows the framework in Figure 2, with specific scales to quantify symptoms, physical and social function, and quality of life. The KCCQ also has a scale to measure self-efficacy, the degree to which patients are confident in managing their heart failure, and changes in symptoms over the past 2 weeks. The KCCQ has been shown to be more sensitive to changes in clinical status than EuroQol (EQ-5D) and its visual analog scale, SF-12, 6-minute walk test, and New York Heart Association (NYHA) functional classification.³¹ While the KCCQ was designed to quantify patient-centered outcomes, it has also been associated with prognosis and costs.^{32–36} Yet, despite the psychometric and prognostic validity of the MLWHF and KCCQ, they are seldom used in clinical trials, observational registries or routine clinical care.²⁰

Generic Health Status Measures

In parallel to disease-specific measures that explicitly seek to quantify how the heart failure syndrome affects patients' lives, researchers have also developed generic measures of health status. These instruments quantify the overall health of patients, including heart failure and all other comorbidities. Common examples include the EuroQol (EQ-5D),³⁷ the Health Utilities Index (HUI),³⁸ the Duke Activity Status Index (DASI),³⁹ and the 12- or 36-item

short form questionnaires (SF-12 or SF-36).⁴⁰ While the advantage of these measures is that they can permit comparisons between heart failure treatments and treatments for other diseases on a common scale, they are less sensitive to important clinical changes than disease-specific measures.⁴¹

Utilities

To perform cost-effectiveness or decision analyses, in which an explicit goal is to combine survival and the quality of that survival, there is a need to distill patients' health status to a single number—a utility—that can be used to determine quality-adjusted life years (QALY) of survival. A utility is thus a quantification of the strength of an individual's preference for any of several possible outcomes of a decision.⁴² This conceptualization of utility has been used widely in economics and public policy, but medical applications have been more limited. In a medical context, a utility rates health status by measuring the value an individual places on his or her current health state. A utility is best measured through an exercise, such as the standard gamble, in which a subject is asked to choose between accepting his or her current state of health and undergoing a procedure that results in either immediate death or perfect health. The risk of death associated with the procedure is varied until the subject is indifferent in a choice between current health state and taking a chance on the procedure. These sequential 'gambles' on immediate death vs. perfect health, yield a risk of dying that can be interpreted as the percent of 'perfect health' that a patient currently has. For example, if one is willing to accept a 15% chance of immediate death to be restored to perfect health, then his/her current utility is 85% of perfect health. A simpler method, the time trade off (TTO), was developed as an alternative to the standard gamble specifically for use in medical applications.^{43, 44} It seems to be more easily interpreted by clinicians and patients, and gives results quantitatively similar to the standard gamble. While utility measures are grounded in theory, they are difficult to assess reliably, and hence are seldom collected in clinical studies. Smaller samples are generally assessed to support modeling approaches that estimate the cost-effectiveness of alternative therapies.

An alternative approach to directly assessing patients' utilities is to use generic health status measures that have been 'mapped' to utility weights. Examples include the EQ-5D, the HUI, DASI, and SF-36. It should be recognized that their scoring is based on a correlation of questionnaire results with standard gamble or TTO results from community-based samples, often those without the disease of interest. While there has been general acceptance that the utility weights assigned to the EQ-5D are reasonable representations of society's 'values' for various health states, other approaches have gained less acceptance and none of these methods represent patients' perspectives of their disease well.

COST

Intensity of resource use is a highly relevant outcome in CER, which aims to determine the relative value of different care approaches. With growing pressure to control health care expenditures, cost endpoints are likely to become increasingly collected and reported. However, the cost of care can only be assessed in the context of the outcomes of that care. Furthermore, the total costs of a therapy, including its impact on downstream costs, is necessary to provide a fair picture of the value of various health care options.⁴⁵ Therefore, the goal must be to compare the relative value of various health care options, by combining total costs with a comprehensive assessment of benefits and risks within the framework of cost-effectiveness analyses.

Unlike the existing decentralized coverage decision process in the United States, most industrialized countries have been more interested in confronting cost-effectiveness from a national level. The National Institute for Health and Clinical Excellence (NICE), a special

authority of the National Health Service in England and Wales, effectively requires cost and utility data to be collected in the drug development process. NICE provides an example of how cost-effectiveness data may be used to compare the relative value of therapies and guide coverage decisions.⁴⁶ Although Americans have generally shown less tolerance for such centralized regulation, there is growing interest in the United States to reduce health care utilization through free choice in response to pricing mechanisms.⁴⁷ Again, CER which collects cost data as an endpoint is essential to help guide such decisions.

Studies that collect data on resource use and cost are difficult to perform and complex to interpret. Because cost studies generally try to sum the totality of resource use in order to assess the relative efficiency of alternative health care interventions, they require consideration of a wide range of endpoints. These include both medical and non-medical costs (loss of work, transportation, etc). Additionally, which resources are included in the endpoint of cost are highly dependent on the perspective chosen. For example, costs realized are different for the patient versus the health system versus the insurance payer versus society, at large. Therefore, before an economic evaluation begins, the perspective of the study must be determined, as it will have important implications for endpoint selection and collection. In interpreting CER with cost data, it should also be recognized that costs assigned to various resources that are collected as endpoints may change rapidly over time. Section 11 in this issue, "Health Economics Issues in CER" addresses many of these issues in far greater detail.

HOSPITALIZATION

Much of the burden of heart failure occurs around acute care for decompensation. There are now over 1 million heart failure hospitalizations annually in the United States alone,⁴ and approximately 70% of total direct costs are attributable to inpatient care.⁴⁸ Hospitalization is typically prompted by severe symptoms. Most patients do not wish to be in the hospital. Iatrogenic complications are common. Recurrent hospitalization is itself a strong marker of subsequent adverse outcomes. Therefore, hospitalization is an event that captures both cost and health status, and is a strong surrogate for subsequent mortality, but does not fit neatly into a single category of clinically relevant endpoints.

Hospitalization can be characterized in a variety of ways. Some studies use first hospitalization as a single time to event measure. For patients who have already been hospitalized for heart failure (a common eligibility criteria or point of capture for many CER studies), readmission may be the endpoint of interest, and these specific rehospitalization events have received particular attention as of late. The endpoint of rehospitalization is paradoxically related to the length of stay of the index hospitalization, because patients with prolonged index hospitalizations have less time at risk for re-hospitalization from the time of index admission. More comprehensive measures of hospitalization may include total number of hospitalizations over some time period or total days of hospitalization, as are needed in cost analyses. In addition, the issues of cause-specific vs. any hospitalization relate to this outcome, as discussed above for mortality. Regardless of the specific hospitalization endpoint chosen, hospitalization should be considered in the context of overall mortality, since patients who do not survive are not at risk for hospitalization.

Hospitalization endpoints may be substantially impacted by social preferences and regional differences in practice patterns. For example, length of stay for heart failure hospitalization in European countries is approximately twice that in the United States.⁴⁹ Additionally, the increased use of "short stay" holding units in emergency departments and the use of intravenous medications in heart failure clinics can confound the definition of hospitalization. With new federal policies designed to reduce readmission rates,

SAFETY

The difference between effectiveness and safety is in some ways a semantic one, as both represent clinical outcomes that are important to patients and are felt to be influenced by the therapeutic option being tested. A single endpoint (e.g. death or hospitalization) may be prevented in some patients yet caused in others through the application of a therapy, such that many endpoints are both effectiveness and safety measures. (Of note, an additional advantage of using all-cause mortality or readmissions is that this endpoint more clearly captures events that are a consequence of adverse outcomes attributable to the safety of the intervention, as opposed to its efficacy.) However, some endpoints (e.g. allergic reaction to a drug, procedural complication) are more clearly undesired side effects of the therapy that are best thought of from a safety perspective. Finally, the nature of safety endpoints may make them less susceptible to treatment selection biases, and therefore more amenable to observational CER approaches.

Given the history of drug development in heart failure, the overall safety profile of new therapies has become an issue of significant concern. In particular, when the focus of a therapy is short-term symptom relief without a long-term benefit on morbidity or mortality (as has commonly been tested in therapies for acute heart failure), establishing that this does not occur at the expense of longer-term safety is critical. In this sense, safety endpoints are effectively non-inferiority measures, requiring specific statistical approaches in order to establish non-inferiority with a prespecified “equivalence boundary”.⁵⁰ Evaluation of safety for new therapies or approaches to care should be guided by an understanding of the therapy’s mechanism as well as by signals from earlier clinical study. This requires testing specific safety hypotheses with the appropriate sample size that reasonably balances the desire to limit the risk for potential post-approval adverse events with the need for efficient pathways for evaluating new therapies.⁵¹ Ultimately, the degree of risk that must be “ruled out” in order to declare a treatment “safe” should be related to the degree and type of benefit (i.e., a drug that improves short-term symptoms only may be held to a higher standard of safety than one that improves mortality).

SURROGATES

A surrogate endpoint is defined by the FDA as “a laboratory measurement or physical sign that is used in therapeutic trials as a substitute for a clinically meaningful endpoint that is a direct measure of how a patient feels, functions, or survives and is expected to predict the effect of the therapy”.⁵² In order to be valid, a surrogate endpoint should meet clearly defined criteria: 1) the surrogate must be in the causal pathway from the intervention to the clinically relevant outcome, as reflected by a strong association between the surrogate and the target; and 2) there must be no important effects of the intervention on the outcome which are not mediated through or captured by the surrogate.⁵¹ Because it is challenging to establish that these criteria are met, regulatory agencies have generally required that new therapies address clinically relevant outcomes before approval, and major policy decisions that favor one therapy over another should also be based on CER data that relates to clinically meaningful findings.

Unfortunately, the history of drug development in heart failure has been marked by recurrent failure of surrogate endpoints to accurately predict clinical outcomes in larger efficacy trials. Improved hemodynamics and increased diuresis have not correlated with more robust clinical endpoints.⁵³ Recent concerns about previous well-accepted surrogate endpoints (glucose, cholesterol, and vascular intimal-medial thickness) have raised the criticism of surrogate endpoints to a new level.^{54–58} Still, surrogate outcomes continue to have an important role in the development and evaluation of health care therapies, because they

often provide a more immediate manifestation of effect and typically allow for shorter and smaller studies. Consequently, early phase clinical studies using surrogate endpoints are necessary to guide the design of larger studies. Some surrogate endpoints deserve particular mention in the design of heart failure CER.

Many surrogates used in heart failure studies relate to estimates of cardiac function and remodeling. For patients with reduced LVEF, improvements in cardiac imaging parameters, particularly left ventricular volumes, has correlated well with long-term hard outcomes. However, ventricular reconstruction surgery does not appear to improve outcomes, even though it directly improves the surrogate.⁵⁹ Hemodynamic parameters and natriuretic peptides share similar properties to cardiac imaging surrogate markers; experience with nesiritide^{60, 61} and levosimendan^{62, 63} show that improvements in B-type natriuretic peptide and pulmonary-capillary wedge pressure do not necessarily correlate with clinically meaningful outcomes.

The interaction between heart failure and the kidney has become a topic of substantially increased interest in recent years.⁶⁴ Worsening renal function in patients with heart failure has been shown to be a powerful predictor of adverse outcomes, and progressive renal failure leads to meaningful clinical events such as worsening fluid retention and initiation of dialysis. Many recent therapies for heart failure have targeted renal preservation as an endpoint.⁶⁵

Yet all commonly used surrogates, such as serum creatinine, LVEF, or B-type natriuretic peptide level, are only important insofar as they illuminate the mechanisms of disease progression and can highlight new pathways for which additional therapeutic interventions might be designed and tested. Ultimately, it is the clinically important outcomes of survival, quality of life, and cost that should guide therapeutic decisions. Given the pitfalls of relying on surrogate endpoints, surrogates should be hypothesis generating and confirmed in studies that directly measure differences in outcomes.

COMBINING ENDPOINTS

There is no single endpoint that accurately captures the totality of the patient experience with heart failure. Thus, substantial interest exists for combining endpoints in order to measure the impact of interventions on the various domains of possible benefit. One method for addressing these issues is the use of multiple endpoints in some type of combination.

Composite endpoints attempt to combine the various aspects of the experience of heart failure into a single integrated measure. Some endpoints can be naturally combined into a single endpoint, if there is a common unit of value to allow for comparison. Examples would include the combination of resource use into a single cost measure or the melding of survival and quality of life into QALYs. However, most endpoints are not so easily condensed. More commonly, a composite endpoint simply combines separate “events” into one category. A frequently used combined endpoint would be “time to death or first hospitalization”. A more nuanced combination would be “days alive and out of the hospital”.⁶⁶ The theoretical advantage of this latter endpoint is that it combines mortality, length of stay of the index hospitalization, and the burden of subsequent hospitalizations into a single endpoint. However, such combinations fundamentally suffer from equating very different outcomes, i.e. death and hospital admission are not equal in the eyes of most patients. Composite endpoints may be more appropriate when either of two endpoints trigger a similar response. For example, many decisions about advanced heart failure therapies are driven by either high risk for death or persistently unfavorable quality of life rather than only one or the other. Therefore, patients at high risk for either endpoint may warrant consideration of aggressive therapies and end-of-life care.

An alternative to combined endpoints is the use of hierarchical endpoints based on ranking of events. This approach is designed to recognize that not all endpoints are of equal importance. In this type of scheme, all patients in a CER cohort are ranked based on a prespecified hierarchy of events. For example, time to death would be ranked at the bottom, then time to hospitalization next, etc. The primary analysis in this type of analysis is a nonparametric comparison of the ranks between those patients receiving one intervention compared to those receiving another. Rank endpoints are felt to be more difficult to interpret and have rarely been used in CER. The advantages and disadvantages of global rank endpoints for trials of mechanical cardiac support devices and acute heart failure have been reviewed in detail.^{67, 68}

Statistically, the use of composites increases the total event rate and therefore may increase statistical power. However, composite endpoints only increase statistical power if the intervention has an effect on multiple aspects of the composite.⁶⁹ Combining endpoints may avoid statistical problems inherent to the individual assessment of non-mortality end points within a population of patients experiencing a high absolute death rate. Patients who die are no longer at risk for other events. If deaths are excluded (for logistic regression) or censored (for time-dependent analyses), an important subpopulation is effectively removed from the analysis. Therefore, most analyses of rehospitalization assess the combined end point of death or rehospitalization for these reasons.

CONSIDERATIONS FOR FUTURE ENDPOINT DESIGN

The discussion above helps provide general principles to guide future CER endpoint selection (Box 1). First and foremost, CER must focus on measures of clinical importance, assessed over a reasonable duration of follow-up. Therefore, greater prospective capture of patient-reported health status and resource use, in addition to mortality, will be essential to realize the full potential of CER. Additionally, standardization and validation of endpoint measures is critical for all types of heart failure research. Creating a consensus on the use of consistent endpoints across studies (e.g. all-cause mortality and rehospitalization, and the MLWHF or KCCQ to measure patient-reported health status outcomes) is critically important so that CER analyses can compare similar outcomes across studies. In fact, the fundamental purpose of CER, to compare various therapies, inherently requires that the same endpoints are being collected in a standardized and reproducible way. The promise of health information technology may bring more detailed clinical data capture, the ability to standardize data capture across health care systems and beyond, and even the ability to prospectively block randomize patients in routine care to various interventions for comparison. Current electronic health record companies have not sought to harmonize their clinical data elements or to support the export of these data to common databases, impeding the potential of CER using these methods.

Box 1

Key considerations for an ideal CER endpoint

1. Measures something of clinical value (i.e. relates to patients feeling better and/or living longer, or saves resources)
2. Simple to obtain
3. Easy to interpret
4. Objective, validated, and accurate
5. Accounts for the totality of patient experience, or combines multiple domains to capture important outcomes

6. Standardized between studies to allow for comparisons

CONCLUSIONS

The field of CER for heart failure continues to evolve, including its assessment of endpoints. Reliance on surrogate endpoints is unacceptable as a means of definitively establishing comparisons of clinical effectiveness. Moving forward, CER will need to focus on measures that clearly reflect clinical effectiveness and safety—not just survival, but also standardized assessments of health status and detailed resource utilization. And it must do so in a standardized way to allow for comparison. This will almost certainly require increased reliance on prospective studies with proactive endpoint capture, preferably in the setting of randomized allocation of the interventions being compared.

References

1. Hunt SA, Abraham WT, Chin MH, Feldman AM, Francis GS, Ganiats TG, Jessup M, Konstam MA, Mancini DM, Michl K, Oates JA, Rahko PS, Silver MA, Stevenson LW, Yancy CW. 2009 focused update incorporated into the ACC/AHA 2005 Guidelines for the Diagnosis and Management of Heart Failure in Adults: a report of the American College of Cardiology Foundation/American Heart Association Task Force on Practice Guidelines: developed in collaboration with the International Society for Heart and Lung Transplantation. *Circulation*. 2009; 119(14):e391–479. [PubMed: 19324966]
2. Dickstein K, Cohen-Solal A, Filippatos G, McMurray JJ, Ponikowski P, Poole-Wilson PA, Stromberg A, van Veldhuisen DJ, Atar D, Hoes AW, Keren A, Mebazaa A, Nieminen M, Priori SG, Swedberg K, Vahanian A, Camm J, De Caterina R, Dean V, Funck-Brentano C, Hellemans I, Kristensen SD, McGregor K, Sechtem U, Silber S, Tendera M, Widimsky P, Zamorano JL. ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure 2008: the Task Force for the Diagnosis and Treatment of Acute and Chronic Heart Failure 2008 of the European Society of Cardiology. Developed in collaboration with the Heart Failure Association of the ESC (HFA) and endorsed by the European Society of Intensive Care Medicine (ESICM). *Eur Heart J*. 2008; 29(19):2388–2442. [PubMed: 18799522]
3. Lindenfeld J, Albert NM, Boehmer JP, Collins SP, Ezekowitz JA, Givertz MM, Katz SD, Klapholz M, Moser DK, Rogers JG, Starling RC, Stevenson WG, Tang WH, Teerlink JR, Walsh MN. HFSA 2010 Comprehensive Heart Failure Practice Guideline. *J Card Fail*. 16(6):e1–194. [PubMed: 20610207]
4. Roger VL, Go AS, Lloyd-Jones DM, Adams RJ, Berry JD, Brown TM, Carnethon MR, Dai S, de Simone G, Ford ES, Fox CS, Fullerton HJ, Gillespie C, Greenlund KJ, Hailpern SM, Heit JA, Ho PM, Howard VJ, Kissela BM, Kittner SJ, Lackland DT, Lichtman JH, Lisabeth LD, Makuc DM, Marcus GM, Marelli A, Matchar DB, McDermott MM, Meigs JB, Moy CS, Mozaffarian D, Mussolino ME, Nichol G, Paynter NP, Rosamond WD, Sorlie PD, Stafford RS, Turan TN, Turner MB, Wong ND, Wylie-Rosett J. Heart Disease and Stroke Statistics--2011 Update: A Report From the American Heart Association. *Circulation*. 2011; 123(6):e240.
5. Allen LA, O'Connor CM. Management of acute decompensated heart failure. *CMAJ*. 2007; 176(6): 797–805. [PubMed: 17353535]
6. Paulus WJ, van Ballegoij JJ. Treatment of heart failure with normal ejection fraction: an inconvenient truth! *J Am Coll Cardiol*. 2010; 55(6):526–537. [PubMed: 20152557]
7. Goodlin SJ, Hauptman PJ, Arnold R, et al. Consensus statement: Palliative and supportive care in advanced heart failure. *J Card Fail*. 2004; 10(3):200–209. [PubMed: 15190529]
8. Stewart S, MacIntyre K, Hole DJ, Capewell S, McMurray JJ. More 'malignant' than cancer? Five-year survival following a first admission for heart failure. *Eur J Heart Fail*. 2001; 3(3):315–322. [PubMed: 11378002]
9. Yusuf S, Negassa A. Choice of clinical outcomes in randomized trials of heart failure therapies: disease-specific or overall outcomes? *Am Heart J*. 2002; 143(1):22–28. [PubMed: 11773908]

10. Hohnloser SH, Kuck KH, Dorian P, Roberts RS, Hampton JR, Hatala R, Fain E, Gent M, Connolly SJ. Prophylactic use of an implantable cardioverter-defibrillator after acute myocardial infarction. *N Engl J Med*. 2004; 351(24):2481–2488. [PubMed: 15590950]
11. O'Connor CM, Miller AH, Konstam MA, Carson P, Haass M, Hauptman PJ, Pena I, Metra M, Sackner-Bernstein JD, Gheorghiade M. Mode of death and cardiovascular re-hospitalization in patients admitted with acute heart failure - results from the EVEREST program. *Eur Heart J*. 2008; (Supp 7):59–60.
12. Carson P, Anand I, O'Connor C, Jaski B, Steinberg J, Lwin A, Lindenfeld J, Ghali J, Barnet JH, Feldman AM, Bristow MR. Mode of death in advanced heart failure: the Comparison of Medical, Pacing, and Defibrillation Therapies in Heart Failure (COMPANION) trial. *J Am Coll Cardiol*. 2005; 46(12):2329–2334. [PubMed: 16360067]
13. Solomon SD, Anavekar N, Skali H, McMurray JJ, Swedberg K, Yusuf S, Granger CB, Michelson EL, Wang D, Pocock S, Pfeffer MA. Influence of ejection fraction on cardiovascular outcomes in a broad spectrum of heart failure patients. *Circulation*. 2005; 112(24):3738–3744. [PubMed: 16330684]
14. Zile MR, Gaasch WH, Anand IS, Haass M, Little WC, Miller AB, Lopez-Sendon J, Teerlink JR, White M, McMurray JJ, Komajda M, McKelvie R, Ptaszynska A, Hetzel SJ, Massie BM, Carson PE. Mode of death in patients with heart failure and a preserved ejection fraction: results from the Irbesartan in Heart Failure With Preserved Ejection Fraction Study (I-Preserve) trial. *Circulation*. 2010; 121(12):1393–1405. [PubMed: 20231531]
15. Mahaffey KW, Harrington RA, Akkerhuis M, Kleiman NS, Berdan LG, Crenshaw BS, Tardiff BE, Granger CB, DeJong I, Bhapkar M, Widimsky P, Corbalan R, Lee KL, Deckers JW, Simoons ML, Topol EJ, Califf RM. Systematic adjudication of myocardial infarction end-points in an international clinical trial. *Curr Control Trials Cardiovasc Med*. 2001; 2(4):180–186. [PubMed: 11806793]
16. Fox CS, Evans JC, Larson MG, Lloyd-Jones DM, O'Donnell CJ, Sorlie PD, Manolio TA, Kannel WB, Levy D. A comparison of death certificate out-of-hospital coronary heart disease death with physician-adjudicated sudden cardiac death. *Am J Cardiol*. 2005; 95(7):856–859. [PubMed: 15781015]
17. Lewis EF, Johnson PA, Johnson W, Collins C, Griffin L, Stevenson LW. Preferences for quality of life or survival expressed by patients with heart failure. *J Heart Lung Transplant*. 2001; 20(9):1016–1024. [PubMed: 11557198]
18. 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(21):1702–1708. [PubMed: 19007689]
19. Stewart GC, Brooks K, Pratibhu PP, Tsang SW, Semigran MJ, Smith CM, Saniuk C, Camuso JM, Fang JC, Mudge GH, Couper GS, Baughman KL, Stevenson LW. Thresholds of physical activity and life expectancy for patients considering destination ventricular assist devices. *J Heart Lung Transplant*. 2009; 28(9):863–869. [PubMed: 19716036]
20. Spertus JA. Evolving applications for patient-centered health status measures. *Circulation*. 2008; 118(20):2103–2110. [PubMed: 19001034]
21. The Criteria Committee of the New York Heart Association. Nomenclature and criteria for diagnosis of diseases of the heart and blood vessels. Boston, MA: Little Brown; 1964.
22. Rafeal C, Briscoe C, Davies J, Ian Whinnett Z, Manistry C, Sutton R, Mayet J, Francis DP. Limitations of the New York Heart Association functional classification system and self-reported walking distances in chronic heart failure. *Heart*. 2007; 93(4):476–482. [PubMed: 17005715]
23. Bennett JA, Riegel B, Bittner V, Nichols J. Validity and reliability of the NYHA classes for measuring research outcomes in patients with cardiac disease. *Heart Lung*. 2002; 31(4):262–270. [PubMed: 12122390]
24. Allen LA, Hernandez AF, O'Connor CM, Felker GM. End points for clinical trials in acute heart failure syndromes. *J Am Coll Cardiol*. 2009; 53(24):2248–2258. [PubMed: 19520247]
25. Pang PS, Cleland JG, Teerlink JR, Collins SP, Lindsell CJ, Sopko G, Peacock WF, Fonarow GC, Aldeen AZ, Kirk JD, Storrow AB, Tavares M, Mebazaa A, Roland E, Massie BM, Maisel AS, Komajda M, Filippatos G, Gheorghiade M. A proposal to standardize dyspnoea measurement in

- clinical trials of acute heart failure syndromes: the need for a uniform approach. *Eur Heart J*. 2008; 29(6):816–824. [PubMed: 18310669]
26. Mahler DA, Weinberg DH, Wells CK, Feinstein AR. The measurement of dyspnea. Contents, interobserver agreement, and physiologic correlates of two new clinical indexes. *Chest*. 1984; 85(6):751–758. [PubMed: 6723384]
 27. Mahler DA, Rosiello RA, Harver A, Lentine T, McGovern JF, Daubenspeck JA. Comparison of clinical dyspnea ratings and psychophysical measurements of respiratory sensation in obstructive airway disease. *Am Rev Respir Dis*. 1987; 135(6):1229–1233. [PubMed: 3592398]
 28. Allen LA, Metra M, Milo-Cotter O, Filippatos G, Reisin LH, Bensimhon DR, Gronda EG, Colombo P, Felker GM, Cas LD, Kremastinos DT, O'Connor CM, Cotter G, Davison BA, Ditttrich HC, Velazquez EJ. Improvements in signs and symptoms during hospitalization for acute heart failure follow different patterns and depend on the measurement scales used: an international, prospective registry to evaluate the evolution of measures of disease severity in acute heart failure (MEASURE-AHF). *J Card Fail*. 2008; 14(9):777–784. [PubMed: 18995183]
 29. Rector TS, Kubo SH, Cohn JN. Validity of the Minnesota Living with Heart Failure questionnaire as a measure of therapeutic response to enalapril or placebo. *Am J Cardiol*. 1993; 71(12):1106–1107. [PubMed: 8475878]
 30. Green CP, Porter CB, Bresnahan DR, Spertus JA. Development and evaluation of the Kansas City Cardiomyopathy Questionnaire: a new health status measure for heart failure. *J Am Coll Cardiol*. 2000; 35(5):1245–1255. [PubMed: 10758967]
 31. Spertus J, Peterson E, Conard MW, Heidenreich PA, Krumholz HM, Jones P, McCullough PA, Pina I, Tooley J, Weintraub WS, Rumsfeld JS. Monitoring clinical changes in patients with heart failure: a comparison of methods. *Am Heart J*. 2005; 150(4):707–715. [PubMed: 16209970]
 32. Soto GE, Jones P, Weintraub WS, Krumholz HM, Spertus JA. Prognostic value of health status in patients with heart failure after acute myocardial infarction. *Circulation*. 2004; 110(5):546–551. [PubMed: 15262843]
 33. Kosiborod M, Soto GE, Jones PG, Krumholz HM, Weintraub WS, Deedwania P, Spertus JA. Identifying heart failure patients at high risk for near-term cardiovascular events with serial health status assessments. *Circulation*. 2007; 115(15):1975–1981. [PubMed: 17420346]
 34. Heidenreich PA, Spertus JA, Jones PG, Weintraub WS, Rumsfeld JS, Rathore SS, Peterson ED, Masoudi FA, Krumholz HM, Havranek EP, Conard MW, Williams RE. Health status identifies heart failure outpatients at risk for hospitalization or death. *J Am Coll Cardiol*. 2006; 47(4):752–756. [PubMed: 16487840]
 35. Chan PS, Soto G, Jones PG, Nallamothu BK, Zhang Z, Weintraub WS, Spertus JA. Patient health status and costs in heart failure: insights from the eplerenone post-acute myocardial infarction heart failure efficacy and survival study (EPHESUS). *Circulation*. 2009; 119(3):398–407. [PubMed: 19139382]
 36. Zuluaga MC, Guallar-Castillon P, Lopez-Garcia E, Banegas JR, Conde-Herrera M, Olcoz-Chiva M, Rodriguez-Pascual C, Rodriguez-Artalejo F. Generic and disease-specific quality of life as a predictor of long-term mortality in heart failure. *Eur J Heart Fail*. 2010; 12(12):1372–1378. [PubMed: 20861126]
 37. [Accessed December 18, 2010] The EQ-5D: An Instrument Designed to Describe and Value Health. www.euroqol.org
 38. [Accessed December 18, 2010] Utility Assessment of Health Related Quality of Life. www.healthutilities.com
 39. Hlatky MA, Boineau ER, Higgenbotham MB, Lee KL, Mark DB, Califf RM, Cobb FR, Pryor DB. A brief self-administered questionnaire to determine function capacity (The Duke Activity Status Index). *Am J Cardiol*. 1989; 64:651–654. [PubMed: 2782256]
 40. [Accessed December 18, 2010] The Short Form 36. www.sf-36.org
 41. Spertus JA, Tooley J, Jones P, Poston C, Mahoney E, Deedwania P, Hurley S, Pitt B, Weintraub WS. Expanding the outcomes in clinical trials of heart failure: the quality of life and economic components of EPHESUS (EPlerenone's neuroHormonal Efficacy and SURvival Study). *Am Heart J*. 2002; 143(4):636–642. [PubMed: 11923800]

42. Von Neumann, J.; Morgenstern, O. *Theory of Games and Economic Behavior*. 1. Princeton, NJ: Princeton University Press; 1944.
43. Torrance GW, Feeny D. Utilities and quality-adjusted life years. *Int J Technol Assess Health Care*. 1989; 5(4):559–575. [PubMed: 2634630]
44. Torrance GW. Utility approach to measuring health-related quality of life. *J Chronic Dis*. 1987; 40(6):593–603. [PubMed: 3298297]
45. Porter ME. What Is Value in Health Care? *New England Journal of Medicine*. 2010; 363(26): 2477–2481. [PubMed: 21142528]
46. Pearson SD, Rawlins MD. Quality, innovation, and value for money: NICE and the British National Health Service. *JAMA*. 2005; 294(20):2618–2622. [PubMed: 16304076]
47. Weinstein MC, Skinner JA. Comparative effectiveness and health care spending--implications for reform. *N Engl J Med*. 2010; 362(5):460–465. [PubMed: 20054039]
48. Liao L, Allen LA, Whellan DJ. Economic burden of heart failure in the elderly. *Pharmacoeconomics*. 2008; 26(6):447–462. [PubMed: 18489197]
49. Blair JE, Zannad F, Konstam MA, Cook T, Traver B, Burnett JC Jr, Grinfeld L, Krasa H, Maggioni AP, Orlandi C, Swedberg K, Udelson JE, Zimmer C, Gheorghiade M. Continental differences in clinical characteristics, management, and outcomes in patients hospitalized with worsening heart failure results from the EVEREST (Efficacy of Vasopressin Antagonism in Heart Failure: Outcome Study with Tolvaptan) program. *J Am Coll Cardiol*. 2008; 52(20):1640–1648. [PubMed: 18992654]
50. Piaggio G, Elbourne DR, Altman DG, Pocock SJ, Evans SJ. Reporting of noninferiority and equivalence randomized trials: an extension of the CONSORT statement. *JAMA*. 2006; 295(10): 1152–1160. [PubMed: 16522836]
51. Reed SD, Anstrom KJ, Seils DM, Califf RM, Schulman KA. Use Of Larger Versus Smaller Drug-Safety Databases Before Regulatory Approval: The TradeOffs. *Health Aff*. 2008; 27(5):w360–370.
52. Prentice RL. Surrogate endpoints in clinical trials: definition and operational criteria. *Stat Med*. 1989; 8(4):431–440. [PubMed: 2727467]
53. Bucher HC, Guyatt GH, Cook DJ, Holbrook A, McAlister FA. Users' guides to the medical literature: XIX. Applying clinical trial results. A. How to use an article measuring the effect of an intervention on surrogate end points. Evidence-Based Medicine Working Group. *JAMA*. 1999; 282(8):771–778. [PubMed: 10463714]
54. Sackner-Bernstein JD, Kowalski M, Fox M, Aaronson K. Short-term risk of death after treatment with nesiritide for decompensated heart failure: a pooled analysis of randomized controlled trials. *JAMA*. 2005; 293(15):1900–1905. [PubMed: 15840865]
55. Gerstein HC, Miller ME, Byington RP, Goff DC Jr, Bigger JT, Buse JB, Cushman WC, Genuth S, Ismail-Beigi F, Grimm RH Jr, Probstfield JL, Simons-Morton DG, Friedewald WT. Effects of intensive glucose lowering in type 2 diabetes. *N Engl J Med*. 2008; 358(24):2545–2559. [PubMed: 18539917]
56. Nissen SE, Wolski K. Effect of rosiglitazone on the risk of myocardial infarction and death from cardiovascular causes. *N Engl J Med*. 2007; 356(24):2457–2471. [PubMed: 17517853]
57. Kastelein JJ, Akdim F, Stroes ES, Zwinderman AH, Bots ML, Stalenhoef AF, Visseren FL, Sijbrands EJ, Trip MD, Stein EA, Gaudet D, Duivenvoorden R, Veltri EP, Marais AD, de Groot E. Simvastatin with or without ezetimibe in familial hypercholesterolemia. *N Engl J Med*. 2008; 358(14):1431–1443. [PubMed: 18376000]
58. Barter PJ, Caulfield M, Eriksson M, Grundy SM, Kastelein JJP, Komajda M, Lopez-Sendon J, Mosca L, Tardif J-C, Waters DD, Shear CL, Revkin JH, Buhr KA, Fisher MR, Tall AR, Brewer B. Effects of Torcetrapib in Patients at High Risk for Coronary Events. *New England Journal of Medicine*. 2007; 357(21):2109–2122. [PubMed: 17984165]
59. Jones RH, Velazquez EJ, Michler RE, Sopko G, Oh JK, O'Connor CM, Hill JA, Menicanti L, Sadowski Z, Desvigne-Nickens P, Rouleau JL, Lee KL. Coronary bypass surgery with or without surgical ventricular reconstruction. *N Engl J Med*. 2009; 360(17):1705–1717. [PubMed: 19329820]

60. Intravenous nesiritide vs nitroglycerin for treatment of decompensated congestive heart failure: a randomized controlled trial. *JAMA*. 2002; 287(12):1531–1540. [PubMed: 11911755]
61. Hernandez AF, O'Connor CM, Starling RC, Reist CJ, Armstrong PW, Dickstein K, Lorenz TJ, Gibler WB, Hasselblad V, Komajda M, Massie B, McMurray JJ, Nieminen M, Rouleau JL, Swedberg K, Califf RM. Rationale and design of the Acute Study of Clinical Effectiveness of Nesiritide in Decompensated Heart Failure Trial (ASCEND-HF). *Am Heart J*. 2009; 157(2):271–277. [PubMed: 19185633]
62. Mebazaa A, Nieminen MS, Packer M, Cohen-Solal A, Kleber FX, Pocock SJ, Thakkar R, Padley RJ, Poder P, Kivikko M. Levosimendan vs dobutamine for patients with acute decompensated heart failure: the SURVIVE Randomized Trial. *JAMA*. 2007; 297(17):1883–1891. [PubMed: 17473298]
63. Packer, M. Randomized multicenter evaluation of intravenous levosimendan efficacy versus placebo in the short-term treatment of decompensated heart failure study (REVIVE-2). Paper presented at: Proceedings of the American Heart Association Scientific Sessions; November 13–16, 2005; Dallas, TX.
64. Ronco C, Haapio M, House AA, Anavekar N, Bellomo R. Cardiorenal syndrome. *J Am Coll Cardiol*. 2008; 52(19):1527–1539. [PubMed: 19007588]
65. Massie BM, O'Connor CM, Metra M, Ponikowski P, Teerlink JR, Cotter G, Weatherley BD, Cleland JG, Givertz MM, Voors A, DeLucca P, Mansoor GA, Salerno CM, Bloomfield DM, Dittrich HC. Rolofylline, an adenosine A1-receptor antagonist, in acute heart failure. *N Engl J Med*. 2010; 363(15):1419–1428. [PubMed: 20925544]
66. Binanay C, Califf RM, Hasselblad V, O'Connor CM, Shah MR, Sopko G, Stevenson LW, Francis GS, Leier CV, Miller LW. Evaluation study of congestive heart failure and pulmonary artery catheterization effectiveness: the ESCAPE trial. *JAMA*. 2005; 294(13):1625–1633. [PubMed: 16204662]
67. Felker GM, Anstrom KJ, Rogers JG. A global ranking approach to end points in trials of mechanical circulatory support devices. *J Card Fail*. 2008; 14(5):368–372. [PubMed: 18514927]
68. Felker GM, Maisel AS. Development of Therapeutics for Heart Failure: A Global Rank End Point for Clinical Trials of Acute Heart Failure. *Circ Heart Fail*. 2010; 3:643–646. [PubMed: 20841546]
69. Ferreira-Gonzalez I, Busse JW, Heels-Ansdell D, Montori VM, Akl EA, Bryant DM, Alonso-Coello P, Alonso J, Worster A, Upadhye S, Jaeschke R, Schunemann HJ, Permyer-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(7597):786. [PubMed: 17403713]

Key Points

- With the increasing availability of therapeutic strategies and the growing complexity of health care delivery, objective evidence of the tangible benefits of different approaches to care is needed, especially for patients with heart failure, a resource-intensive disease.
- Few well-proven therapies exist for patients with acute decompensation or for patients with normal LVEF.
- Comparative effectiveness research (CER) offers an important avenue for making progress in the field; however, CER, like any well-designed research program, requires articulation of clinically important outcomes to be compared.
- For patients with heart failure, endpoint measures that capture the totality of potential benefits and risks for alternative therapeutic approaches must be developed.
- For one therapeutic approach to be considered superior to another, it must make patients live longer, make them feel better, or save money without adversely affecting the other two goals.
- Importantly, these outcomes must be measured directly and surrogates should be avoided, even if such surrogates appear to be associated with clinically meaningful, patient-centered outcomes.

Figure 1. Clinical course of heart failure, illustrating some of the events that may be captured as endpoints. Adapted from Goodlin SJ, Hauptman PJ, Arnold R, et al. Consensus statement: Palliative and supportive care in advanced heart failure. *J Card Fail.* 2004;10(3):200–209

Figure 2.
A conceptual model of health status

Table 1

Categories of potential endpoints for comparative effectiveness research in heart failure

Category	Specific Endpoints	Advantages	Disadvantages
Vital Status (Quantity of life)	Mortality <ul style="list-style-type: none"> • All-cause • Cardiovascular • Pump failure • Sudden death 	The gold standard; highly objective, easy to obtain accurately on patients over long-term follow up	N likely needs to be large and follow up long; therapies which improve many patients, especially with advanced heart failure, may value trade quantity for quality of life
Health Status (Quality of Life)	Health status questionnaires <ul style="list-style-type: none"> • Disease-specific • Generic Utilities NYHA functional classification Symptom scales	By definition reflects how patients feel, making them clinically relevant	Must be obtained directly from patient, prospectively; somewhat time consuming questionnaires
Cost (Resource use)	Inflation adjusted dollars <ul style="list-style-type: none"> • Direct v. indirect • Fixed v. variable 	An objective outcome with patient and societal importance; the numerator in cost-effectiveness analyses used to determining the relative value of care, which is a major goal of CER	Must interpret in context of therapy effectiveness; (cost-effectiveness/cost-utility measures are more meaningful than cost alone); it is difficult to directly capture all costs; often requires multiple assumptions that are not well characterized; critically dependent on the perspective taken (i.e., patient, payer, society)
Mixed (Hospitalization)	Index admission Length of stay Readmissions Total days in hospital	Important clinical measure of cost, but also closely linked to health status and predictor or mortality; easy to measure; relatively objective	Variable between providers, institutions, and countries as standards for admission and dischargability do not exist; surrogate for health status
Surrogates	Vital signs Blood tests Imaging studies	Typically objective; short follow up needed; continuous variables; allow for smaller shorter early phase studies	Frequently surrogate measures do not relate to clinically relevant events
Composites	Combined Ranked	May increase power; may help capture a wider range of important outcome domains; accounts for competing events (i.e. death) rather than censoring such events	More difficult to interpret; least important component often drives result
Safety	Medication side effects Procedural complications	May be less susceptible to treatment selection biases and confounding	May require non-inferiority statistical approaches to say that one approach is no more dangerous than another; some side effects are rare and occur far out from start of intervention, requiring alternatives to RCT for detection (phase IV surveillance)

NYHA = New York Heart Association; CER = comparative effectiveness research

Table 2

Examples of health-related quality of life instruments that may be used in CER

	Description	Scoring	Advantages	Disadvantages
<i>Utilities</i>	Quantification of strength of preference for current health status versus perfect health status with risk of early death		<i>Can be readily combined with survival; necessary for cost-effectiveness analyses</i>	<i>Relatively difficult to obtain from patients</i>
Standard Gamble ⁴²		Percent risk of immediate death a patient is willing to accept for return to perfect health	Standard method for determination of utilities	
Time Trade Off ⁴³		Percent of remaining lifespan a person is willing to trade away for perfect health	More easily interpreted by clinicians and patients	
<i>Generic Health Status Measures</i>	<i>Assessments of patients' symptoms, functioning, and quality of life</i>		<i>Easily compared across studies; map to utilities</i>	<i>Insensitive to important clinical changes in patients with heart failure</i>
EQ-5D ³⁷	5 dimensions <ul style="list-style-type: none"> • Mobility • Self-care • Usual activities • Pain/discomfort • Anxiety/depression 	Each item: no, some, extreme; score -1 to 1		
HUI ³⁸	Multiple dimensions depending on version (vision, hearing, speech, ambulation, dexterity, emotion, cognition, pain, cognitive, self-care, fertility)	Score 0–1		
SF-36 ⁴⁰	8 dimensions <ul style="list-style-type: none"> • Physical functioning • Role limitations physical • Bodily pain • General health • Vitality • Mental health • Role limitations emotional • Social functioning 	Each item: yes or no; score 0–100		
<i>Disease-specific Health Status Measures</i>	<i>Assessments of patients' symptoms, functioning, and quality of life</i>		<i>Sensitive to important clinical changes</i>	<i>Have not been mapped to utilities</i>
MLWHF ²⁹	21-items <ul style="list-style-type: none"> • Physical • Mental 	105-0 (higher scores worse)		
KCCQ ³⁰	23-items in 6 domains:	0–100 (higher scores better)		

	Description	Scoring	Advantages	Disadvantages
	<ul style="list-style-type: none">• Physical limitation• Symptoms• Quality of life• Social interference• Self-efficacy• Overall summary			

EQ-5D = EuroQol; HUI = Health Utilities Index; SF = short form; MLWHF= Minnesota Living With Heart Failure Questionnaire; KCCQ = Kansas City Cardiomyopathy Questionnaire