End Points for Clinical Trials in Acute Heart Failure Syndromes

Larry A. Allen, MD, MHS,* Adrian F. Hernandez, MD, MHS,† Christopher M. O’Connor, MD,† G. Michael Felker, MD, MHS†

Aurora, Colorado; and Durham, North Carolina

Acute heart failure syndromes (AHFS) have a major cause of morbidity and mortality, in part because the development of new therapies for these disorders has been marked by frequent failure and little success. The heterogeneity of current approaches to AHFS drug development, particularly with regard to end points, remains a major potential barrier to progress in the field. End points involving hemodynamic status, biomarkers, symptoms, hospital stay, end organ function, and mortality have all been employed either alone or in combination in recent randomized clinical trials in AHFS. In this review, we will discuss the various end point domains from both a clinical and a statistical perspective, summarize the wide variety of end points used in completed and ongoing AHFS studies, and suggest steps for greater standardization of end points across AHFS trials. (J Am Coll Cardiol 2009;53:2248–58) © 2009 by the American College of Cardiology Foundation

Acute heart failure syndromes (AHFS) are a family of related clinical entities characterized by new or worsening signs or symptoms of heart failure leading to hospital stay. Taken together these disorders represent a major public health problem, with a high degree of morbidity and a rapidly increasing prevalence (1). The adverse impact of AHFS on public health has been compounded by the failure to develop new and effective therapies for this syndrome (2). Consequently, guideline recommendations are sparse and largely driven by expert opinion (3,4). Despite significant advances in the care of patients with stable chronic heart failure, neither routine therapy for patients with an episode of AHFS nor the subsequent risk of death have changed significantly over the past 30 years (5,6).

Why has there been so little progress in AHFS therapy? Some of the explanation relates to patient heterogeneity, as exemplified by the plural term “acute heart failure syndromes.” This group of disorders includes patients with diverse presentations and pathophysiology, ranging from patients presenting suddenly with hypertension and normal or near normal ejection fraction to those with subacute presentations with advanced systolic dysfunction and low output states. Given that it is unlikely that the same therapy would be efficacious in such varied patient populations, it is not surprising that “one-size-fits-all” attempts at developing new therapies have not met with success. Additionally, understanding of underlying pathophysiologic mechanisms for AHFS remains limited, and a framework for classification of these syndromes that incorporates both clinical and pathophysiologic differences is lacking. Finally, as we will focus on in this review, the methodology of clinical trials in this area has remained poorly developed (7). There has been little consensus in the research or regulatory community about a variety of aspects of clinical trial design in AHFS, resulting in a heterogeneous approach to drug development. Nowhere is this more apparent than in the lack of agreement about the most appropriate end points for clinical trials in AHFS. This is in sharp contrast to the field of acute coronary syndromes, where there has been broad consensus around the optimal efficacy and longer-term survival (Table 1). Greater standardization of end points across studies is needed for therapies to be easily compared and prioritized by clinicians and payers. Additionally, better validated surrogates for both efficacy and
safety are needed to help speed the development of new agents by reliably recognizing efficacy and safety issues earlier in the development process. In this review, we will summarize end points used in previous and ongoing AHFS studies, describe the current regulatory environment for the approval of AHFS therapies, and discuss the key issues for the development of more robust end points moving forward.

**History of Drug Development in AHFS**

The evaluation and requirements for approval of AHFS therapies have evolved significantly over the years. Initial therapies for AHFS such as loop diuretic agents were accepted into practice largely on clinical experience in combination with pharmacokinetic and pharmacodynamic data. Milrinone was approved for AHFS by the U.S. Food and Drug Administration (FDA) in 1988 on the basis of its hemodynamic effects rather than on hard clinical end points. No additional agents were approved specifically for use in AHFS until 2001, when nesiritide was granted approval by the FDA on the basis of short-term improvements in both hemodynamic status and patient-reported dyspnea (8). Subsequently, there have been substantial concerns raised about the safety and efficacy of both milrinone (9) and nesiritide (10,11), which have led to increased skepticism among both the clinical and the regulatory community about the end points on which to base conclusions about drug efficacy and safety in AHFS. Since the approval of nesiritide by the FDA in 2001, multiple novel therapies for AHFS have been tested in a variety of clinical trials, almost all of which used differing end points and which have resulted in a variety of interpretations (Table 2) (12). This uncertainty as to how to interpret the evidence has led to AHFS therapies that are approved for use in the U.S. but not in Europe (nesiritide) and, conversely, drugs approved for use in Europe but not in the U.S. (levosimendan). In the following text, we will summarize the advantages and disadvantages of various types of end points that have been employed in clinical trials of AHFS.

**Clinical End Points**

The value of a medical intervention is ultimately defined by how it affects the lives of patients. In general, the FDA has mandated that for a therapy to be approved it has to make patients feel better and/or live longer (13). Consequently, pivotal phase III trials of new therapies must demonstrate clinically relevant improvement in a clinical end point (as distinct from a surrogate end point) to justify regulatory approval and clinical use.

**Mortality.** Mortality is obviously an important end point in any clinical syndrome associated with substantial risk of death (such as AHFS). Mortality is objective and easy to assess, and its clinical relevance is self-evident. Controversy remains about whether it is preferable to measure all-cause or disease-specific mortality (14). All-cause mortality has the advantage of a higher overall event rate than disease-specific mortality as well as requiring no adjudication as to cause of death. Importantly, however, all-cause mortality will include events that are unlikely to be responsive to the therapy being tested, which will tend to diminish the overall power of the analysis. Disease-specific event rates related to the specific mechanism of action of the therapy (e.g., sudden death in defibrillator trials) are much more likely to demonstrate a treatment effect, assuming that the mechanism of action of the therapy is reasonably well understood. The use of disease-specific event rates generally necessitates the use of clinical event committees to adjudicate cause of death. Although the process of adjudication might be challenging due to frequent comorbidity in patients with heart failure or a lack of adequate documentation of the event, the use of some form of clinical event adjudication has become standard in major cardiovascular trials (15). Recent data from the EVEREST (Efficacy of Vasopressin Antagonism in Heart Failure Outcome Study with Tolvaptan) trial found that most post-discharge deaths in patients with AHFS and systolic dysfunction are cardiovascular in nature, suggesting that use of disease-specific mortality might be less critical in this population (16). Although data on cause of death in patients with AHFS and normal ejection fraction are lacking, experience indicates that a higher proportion of mortality in such patients might be noncardiac (17), suggesting that disease-specific mortality might be a superior choice for AHFS studies focused on this population.

Despite its obvious appeal as a clinical end point, mortality has not been used as the sole primary end point in any large clinical trials of AHFS therapies. Although AHFS is a highly morbid condition with in-hospital mortality rates on the order of 4% to 7% and 30- to 60-day mortality rates of 10% to 12% (3,4), trial designs have not targeted mortality alone, because of the heterogeneity of the population, the high prevalence of contributing comorbidities, and the recurrent failure of prior AHFS studies to positively impact survival. Additionally, it is uncertain the extent to which short-term AHFS therapies given during an index hospital stay can impact post-discharge mortality. For a short-term AHFS therapy to affect mortality rates, it would need to: 1) significantly diminish in-hospital mortality; 2) alter the fundamental natural history of the AHFS syndrome in a way that leads to a reduction in subsequent clinical events; or 3) facilitate the introduction or up-titration of chronic heart failure therapies known to improve survival. For example, acute reperfusion therapy for myocardial infarction conveys a long-term survival benefit by arresting the central pathophysiologic mechanism (ongoing
myocardial damage due to thrombosis in a coronary artery), with subsequent prevention of disease progression. Whether such a paradigm can be translated to AHFS therapies is unknown, given that no over-arching fundamental mechanism has been identified for this heterogeneous family of disorders.

Heart failure hospital stay. Much of the morbidity burden of heart failure is the result of inpatient care, with over 1 million heart failure hospital stays annually in the U.S. alone (1). Hospital stay for AHFS is associated with severe symptoms, diminished quality of life, worsened post-discharge prognosis, and substantial cost. For patients institutionalized with AHFS, both the LOS for the index hospital stay (median of 6 days in the U.S.) and the risk of repeat hospital stay (30% at 60 days) are substantial (18,19). Thus, addressing the morbidity of AHFS by preventing repeat hospital stay and limiting LOS is an obvious therapeutic objective in AHFS.

Unlike mortality end points, hospital stay end points might be substantially impacted by social preferences and

| Table 1 Categories of Potential End Points for Large Clinical Trials of AHFS Interventions |
|-----------------------------------------------|-----------------------------------------------|-----------------------------------------------|-----------------------------------------------|
| Category                                      | Specific End Points                            | Advantages                                    | Disadvantages                                | Approximate Sample Size |
| Objective clinical outcomes                   | “Hard” measures that quantify morbidity and mortality and generally drive resource use | Require large sample sizes; might not necessarily capture short-term but clinically relevant improvements in patient QOL | Generally more difficult to capture; wide variety of reasons for such visits; generally of less significance than a hospital stay | Large |
| Mortality                                     | Ideal standard for efficacy, highly objective, easiest to obtain accurately on patients over long-term follow-up | Number of patients needs to be large and follow-up long; therapies that improve QOL/symptoms but not mortality are potentially valuable | Highly variable among providers, institutions, and countries; standards for dischargability do not exist | $10^2$-$10^3$ |
| LOS                                           | Important clinical measure for patient QOL and societal resource use | Follow-up period for event might be affected by differences in LOS | Includes LOS and is thus subject to the same limitations | $10^2$-$10^3$ |
| Repeat hospital stay rates                    | Important clinical measure for patient QOL and societal resource use | | | $10^2$-$10^3$ |
| Total days of hospital stay                   | Important clinical measure for patient QOL and societal resource use | | | $10^2$-$10^3$ |
| Urgent care visits                            | Frequent, clinically meaningful | | | $10^2$-$10^3$ |
| Worsening heart failure/need for rescue therapy | | | | $10^2$-$10^3$ |
| Subjective clinical measures                  | By definition relate to how patients feel, theoretically making them clinically relevant | Often highly subjective; generally improve rapidly after admission regardless of therapy | Lacks standardization; largely improves regardless of intervention | $10^2$-$10^3$ |
| Dyspnea (VAS, Likert)                         | Accepted | | | $10^2$-$10^3$ |
| General well-being/global status              | Might represent more aspects of the AHFS experience than dyspnea alone | Less standard and less defined than dyspnea | Relatively complex and time-consuming questionnaires | $10^2$-$10^3$ |
| Health status (KCCQ, MLHF, EQ-5D)            | Validated in chronic heart failure | | | $10^2$-$10^3$ |
| Surrogates                                    | Typically objective; continuous variables increase power | Question clinical relevance | | |
| Hemodynamic status (PCWP)                     | Objective, correlate with symptoms | Has not correlated with important clinical measures | | $10^1$-$10^2$ |
| Weight loss                                   | Noninvasive, global measure of fluid status | Might not correlate highly with symptoms or discharge; reliability is questionable | | $10^1$-$10^2$ |
| BNP                                           | Objective, correlate with symptoms | Changes in BNP as effected by different therapies have not been validated to correlate with important clinical measures | | $10^1$-$10^2$ |
| Creatinine (GFR)                              | Creatinine is a strong predictor of outcome | Unknown whether interventions that change creatinine correlate with clinical outcomes | | $10^1$-$10^2$ |
| Composites                                    | Combined | Might increase power | More difficult to interpret; least important component often drives result | $10^2$-$10^3$ |
| Ranked                                        | Might increase power; graded importance of the various components | Not commonly used, less-familiar; statistically more complex | | $10^2$-$10^3$ |
| Cost                                          | An objective outcome with patient and societal importance | Often requires assumptions about multiple variables leading to wide margins of confidence | | $10^2$-$10^3$ |
| Safety                                        | Given track record of some AHFS therapies that improve short-term measures but might worsen later outcomes, measures of longer-term events take on a key role | Assumes noninferiority statistical properties that tend to greatly increase required sample sizes; as such most safety analyses are ad hoc | | $10^4$-$10^5$ |

Note: AHFS = acute heart failure syndromes; BNP = B-type natriuretic peptide; EQ-5D = EuroQol 5-Dimensional health related quality of life descriptive system; GFR = glomerular filtration rate; KCCQ = Kansas City Cardiomyopathy Questionnaire; LOS = length of stay; MLHF = Minnesota Living with Heart Failure questionnaire; PCWP = post-capillary wedge pressure; QOL = quality of life; VAS = visual analog scale.
Worsening heart failure/need for rescue therapy. Traditional "time to event" analyses using Cox proportional hazards models, which have become standard in chronic heart failure trials, might be less relevant when the follow-up period is relatively short (months rather than years). Time-to-event methods also censor patients after the initial event, thus discounting the clinical burden of multiple or prolonged hospital stays. Additionally, the end point of "repeat hospital stay" is paradoxically related to the LOS of the index hospital stay, because patients with prolonged index hospital stays have less time at risk for repeat hospital stay. Similarly, hospital stay must be considered in the context of overall mortality, because patients who do not survive are not at risk for repeat hospital stay. A composite of death or heart failure hospital stay has been used as the co-primary end point for the recently completed EVEREST study and the ongoing ASCEND–HF (Acute Study of Clinical Effectiveness of Nesiritide in Decompensated Heart Failure) study (22,23).

Another approach that might more completely capture the burden of mortality and hospital stay during the follow-up period is the end point of "days alive and out of the hospital" (9,24). The theoretical advantage of this end point is that it combines mortality, LOS of the index hospital stay, and the burden of subsequent hospital stays into a single end point. However, when index LOS is especially long (which might be of particular concern in studies enrolling a majority of patients outside North America), it might lead to decreased power. Statistical modeling performed for the purposes of designing the ASCEND–HF study suggests that, for interventions that make an impact on LOS for the index hospital stay, "days alive and out of hospital" end points have greater power than the composite of death and repeat hospital stay. Conversely, for interventions without an impact on initial LOS, the composite of death and repeat hospital stay is the more powerful end point (23).

Worsening heart failure/need for rescue therapy. Traditional end points for AHFS studies have tended to focus on short-term symptomatic status (discussed in detail in the following text) or post–discharge outcomes such as mortality or repeat hospital stay. This dichotomy notably neglects a critical period of the AHFS episode, the time between initial stabilization (24 to 48 h) and eventual hospital discharge (Fig. 1). Recognition that commonly used end points do not reflect this important time period has led to the development of "worsening heart failure" during therapy as a novel outcome measure in AHFS. Although there is no consensus definition of worsening heart failure, typically it is defined as either failure to improve (persistent signs and symptoms of heart failure despite therapy) or worsening signs and symptoms of heart failure despite therapy. One component of the worsening heart failure end point is the requirement for "rescue therapy" (i.e., the need to initiate or intensify intravenous therapy [such as inotropes or intravenous vasoactive agents] or implement mechanical cardiac or ventilatory support). Although the need for such rescue therapy makes intuitive sense as part of an end point designed to capture lack of clinical improvement, guidance from European Medicines Agency (EMEA) suggests that they do not consider this an appropriate component of an efficacy end point (25). Multiple recent trials have used worsening heart failure end points as either a component of the primary end point or an important secondary end point (26–28).

Symptom measures. Acute heart failure syndrome is a highly symptomatic disorder, and faster or more complete resolution of symptoms is an important clinical goal for AHFS therapy. Patient–reported symptom measures have become the most commonly used end points in AHFS trials, either alone or in combination with others measures, consistent with the fundamental importance of helping patients feel better or live longer.

Dyspnea has emerged as the current standard for patient-reported clinical end point measures (29). Dyspnea is the most common presenting symptom for patients with AHFS, and hospital discharge is often dictated by resolution in dyspnea. Most contemporary phase III AHFS trials have used some measure of short-term changes in dyspnea as a key end point (23,27,30) (Table 2).

Despite the apparent ascendance of dyspnea as a key measure of efficacy in AHFS trials, there are significant issues regarding dyspnea as an end point. No validated instrument for dyspnea assessment currently exists that is accurate, reliable, reproducible between observers, and sensitive to important changes in dyspnea (31). This has led to the use of an assortment of poorly validated instruments for assessing dyspnea, including Likert scales, visual analog scales, and more complicated measures that incorporate patient effort and time variables, such as the Baseline Dyspnea Index (32) and Transient Dyspnea Index (33). In a prospective registry of patients with AHFS, patterns of dyspnea resolution were significantly affected by choice of response instrument (34). Likert measures of dyspnea initially improved rapidly with no significant improvement thereafter, whereas visual analog scale measurement of dyspnea improved continually throughout hospital stay. Recently, a proposal for greater standardization of dyspnea measurement in AHFS trials suggested sequential evaluation of dyspnea with progressively greater provocation (lying flat, walking, and so forth) (31). Although intriguing, such an instrument might be somewhat laborious to conduct, and it requires further validation before it should be widely implemented.
## Selected Large Phase 2 and 3 AHFS Trials and Their End Points

<table>
<thead>
<tr>
<th>Study Acronym (Ref. #)</th>
<th>Year of Publication</th>
<th>Intervention</th>
<th>Primary End Point(s)</th>
<th>Key Secondary End Points</th>
<th>Long-Term Clinical Outcomes</th>
</tr>
</thead>
<tbody>
<tr>
<td>VMAC (8)</td>
<td>2002</td>
<td>Nesiritide infusion 48 h vs. nitroglycerine infusion vs. placebo</td>
<td>Coprimary: 1. Change in PCWP at 3 h 2. Change in dyspnea (Likert) at 3 h</td>
<td>PCWP at 24 h, dyspnea at 24 and 48 h, global clinical status</td>
<td>Tertiary end points: 48 h hypotension, headaches; 30-day mortality, myocardial infarction, readmission, and renal dysfunction; 6-month mortality (not powered for definitive analysis)</td>
</tr>
<tr>
<td>OPTIME (9)</td>
<td>2002</td>
<td>Milrinone infusion 48 h vs. placebo</td>
<td>Cumulative days of hospital stay for cardiovascular cause or days dead within 60 days after randomization</td>
<td>Proportion of cases failing therapy due to adverse events or worsening heart failure (sustained SBP &lt; 80 mm Hg, myocardial ischemia, arrhythmias, persistent CHF, inadequate diuresis, organ hypoperfusion), heart failure score, global health (VAS)</td>
<td>Primary end point</td>
</tr>
<tr>
<td>ESCAPE (24)</td>
<td>2005</td>
<td>Therapy guided by pulmonary artery catheter plus clinical assessment vs. clinical assessment alone</td>
<td>Days alive and out of hospital during the first 6 months</td>
<td>Adverse events related to catheter use, 6-min walk duration, QOL via time trade-off, and MLHF</td>
<td>Primary end point</td>
</tr>
<tr>
<td>VERITAS (27)</td>
<td>2007</td>
<td>Tezosentan infusion 24–72 h vs. placebo</td>
<td>Co-primary: 1. Change in dyspnea (at 3, 6, and 24 h with VAS 0–100) over 24 h (area under the curve) 2. Death or worsening heart failure (pulmonary edema, shock, new or ↑ intravenous therapy, mechanical cardiac or pulmonary support, renal replacement therapy) at 7 days</td>
<td>Death or major cardiovascular events at 30 days; improved hemodynamic measures over 24 h; LOS; days hospitalized within 30 days; 6-month mortality</td>
<td>Secondary end points</td>
</tr>
<tr>
<td>SURVIVE (46)</td>
<td>2007</td>
<td>Levosimendan infusion vs. dobutamine infusion as long as clinically indicated in AHFS patients requiring inotropic support</td>
<td>All-cause mortality at 180 days</td>
<td>All-cause mortality at 31 days; days alive or out of hospital at 180 days; cardiovascular mortality at 180 days; change in BNP level at 24 h; dyspnea at 24 h; patient-assessed global assessment at 24 h</td>
<td>Primary and secondary end points</td>
</tr>
<tr>
<td>REVIVE-2 (26)</td>
<td>Not yet published (presented 2009)</td>
<td>Levosimendan infusion vs. placebo in hemodynamically stable AHFS patients</td>
<td>Composite of clinical signs and symptoms of heart failure over 5 days expressed as 3-stage end point: 1. Better (moderately or markedly improved global assessment at 6 h, 24 h, and 5 days with no worsening) 2. Same 3. Worse (death from any cause, persistent or worsening heart failure requiring intravenous diuretic agents, vasodilators, or inotropes at any time; or moderately or markedly worse patient global assessment at 6 h, 24 h, or 5 days)</td>
<td>Change in BNP; mortality at 90 days</td>
<td>Secondary end point</td>
</tr>
<tr>
<td>EVEREST (22,35)</td>
<td>2007</td>
<td>Tolvaptan vs. placebo up to 112 weeks</td>
<td>Short-term composite: changes in global clinical status (by VAS) and body weight at day 7 or discharge Long-term dual end points: 1. All-cause mortality (superiority and noninferiority) 2. Cardiovascular death or heart failure hospital stay (superiority only)</td>
<td>Composite components in isolation at days 1 and 7 or discharge; dyspnea at day 1; peripheral edema at day 7 or discharge; KCCQ at 1 week and 6 months; body weight; changes in serum sodium</td>
<td>Main safety end point</td>
</tr>
<tr>
<td>ASCEND-HF (23)</td>
<td>Enrolling</td>
<td>Nesiritide infusion vs. placebo</td>
<td>Coprimary: 1. Composite of all-cause mortality and heart failure repeat hospital stay through 30 days 2. Dyspnea at 6 and 24 h</td>
<td>Overall well-being (Likert) 6 and 24 h; days alive and outside of hospital within 30 days</td>
<td>Primary end point</td>
</tr>
</tbody>
</table>

Continued on next page
Perhaps most problematic for the use of dyspnea as a primary end point is the possibility of relatively rapid improvement irrespective of therapy for AHFS. In the placebo arms of many recent AHFS studies, there was substantial relief of dyspnea within 24 to 48 h with standard therapy alone (27,30,35). More recently, studies that required a more objective measure of disease severity for patient enrollment (such as natriuretic peptide levels) have not shown the same degree of dyspnea improvement in the placebo group, suggesting that more severely ill patients might have more unresolved dyspnea that could serve as a target for dyspnea-reducing therapy (28). In any case, for an incremental improvement in dyspnea to be considered a significant therapeutic advance, it would need to be relatively rapid, of substantial magnitude, and sustained beyond the initial few hours of therapy.

Measures of patient-reported global health status, also referred to as general well-being, have also been incorporated into AHFS studies in a fashion similar to dyspnea (35). Global health status measures have the potential to better summarize the overall subjective state of being that a patient is experiencing, thus capturing other domains of AHFS that might not be reflected by dyspnea measures. The downside is a potential decrease in power due to inclusion of symptoms that are not affected by the intervention. In general, data suggest that measures of general well-being and those of dyspnea are highly concordant in patients with AHFS, indicating that the choice of symptom domain might not be critical (34).

**Surrogate End Points**

A surrogate end point 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 end point that is a direct measure of how a patient feels, functions, or survives and is expected to predict the effect of the therapy” (36). To be valid, a surrogate end point 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 that are not mediated through or captured by the surrogate (37). Because it is challenging to establish that these criteria are met, surrogate measures are generally not accepted as proof of efficacy but rather as a signal of effect. Currently, there are no widely accepted surrogate end points in heart failure. Regulatory agencies have generally required that new therapies address clinically relevant outcomes before approval. Recent concerns about well-established surrogate end points (cholesterol and glucose reduction) support such a policy (38–40). Still, surrogate outcomes continue to have an important role in the development of new therapies, because they often provide a more immediate manifestation of effect and typically allow for shorter and smaller trials in early phase development. Consequently, early phase clinical studies designed to provide “proof of concept” or to select dosing for larger studies typically employ surrogate end points.

Unfortunately, the history of drug development in heart failure has been marked by the frequent failure of surrogate end points to accurately predict clinical outcomes in larger efficacy trials. In the following text, we review possible surrogate end points in the development of new AHFS therapies.

**Hemodynamic status.** Until recently, changes in hemodynamic status have been the primary focus of therapies for AHFS. Early trials of milrinone focused almost entirely on effects on filling pressures and cardiac output. In the VMAC (Vasodilation in the Management of Acute Congestive Heart Failure) study, post-capillary wedge pressure was 1 of the 2 primary end points ultimately used for FDA approval of nesiritide (8). Despite the hemodynamic benefits of nesiritide and milrinone, significant questions remained about the safety and efficacy of both agents (9,11).

---

**Table 2 Continued**

<table>
<thead>
<tr>
<th>Study Acronym (Ref. #)</th>
<th>Year of Publication</th>
<th>Intervention</th>
<th>Primary End Point(s)</th>
<th>Key Secondary End Points</th>
<th>Long-Term Clinical Outcomes</th>
</tr>
</thead>
<tbody>
<tr>
<td>PROTECT I and II (49)</td>
<td>Enrolling</td>
<td>Rolofylline infusion vs. placebo</td>
<td>Composite of clinical signs and symptoms of heart failure over 7 days expressed as 3-stage end point: 1. Better (moderately or markedly improved global assessment at 24 and 48 h with no worsening) 2. Same 3. Worse (death from any cause, persistent or worsening heart failure through day 7, or creatinine increase ≥0.3 mg/dl at 7 and 14 days)</td>
<td>Safety; within trial costs</td>
<td>Secondary end point</td>
</tr>
</tbody>
</table>

ASCEND-HF = Acute Study of Clinical Effectiveness of Nesiritide in Decompensated Heart Failure; CHF = congestive heart failure; ESCAPE = Evaluation Study of Congestive Heart Failure and Pulmonary Artery Catheterization Effectiveness; EVEREST = Efficacy of Vasopressin Antagonism in Heart Failure Outcome Study with Tolvaptan; OPTIME = Outcomes of a Prospective Trial of Intravenous Milrinone for Exacerbations of Chronic Heart Failure; PROTECT = Study of the Selective A1 Adenosine Receptor Antagonist KW-3902 (rolofylline) for Patients Hospitalized With Acute HF and Volume Overload to Assess Treatment Effect on Congestion and Renal Function; REVIVE-2 = Second Randomized Multicenter Evaluation of Intravenous Levosimendan Efficacy; SBP = systolic blood pressure; SURVIVE = Survival of Patients with Acute Heart Failure in Need of Intravenous Inotropic Support; VERITAS = Value of Endothelin Receptor Inhibition with Tezosentan in Acute Heart Failure Study; VMAC = Vasodilation in the Management of Acute CHF; other abbreviations as in Table 1.
Additionally, hemodynamic parameters have not necessarily been successful in guiding dose selection in the development of new AHFS therapies. As an example, the Randomized Intravenous Tezosentan drug development program studied the endothelin antagonist at doses of 50 to 100 mg/h, on the basis of the dose range that produced the most significant hemodynamic effect in early phase I drug development (41). Overall, this phase II program demonstrated uncertain clinical benefit but a clear signal for greater adverse events in the higher doses, suggesting that the dose range selected on the basis of maximizing hemodynamic benefit was too high for a favorable risk-benefit ratio. Subsequently, the phase III VERITAS (Value of Endothelin Receptor Inhibition with Tezosentan in Acute Heart Failure Study) used a dose of 1 mg/h (27). Thus, despite extensive early phase hemodynamic studies, the “best dose” of tezosentan remained unknown (at a 50-fold range) well into phase III development.

**Fluid balance/weight.** Congestion is widely recognized as playing a central role in the pathophysiology of AHFS, and therapy aimed at reducing fluid congestion (primarily loop diuretic agents) is prescribed to 90% of patients admitted for AHFS in the U.S. (42). Recently, a variety of new therapeutic approaches have been developed to address congestion in AHFS, including vasopressin antagonists, adenosine antagonists, and ultrafiltration. This has led to a substantial interest in the use of change in body weight as a potential objective surrogate end point reflecting decongestion. In the largest trial of ultrafiltration, weight loss at 48 h was the co-primary end point (along with dyspnea) (30). Similarly, in the EVEREST trial the primary end point was a composite including changes in body weight (along with patient-reported global clinical status) (35). In both of these trials, it was the change in fluid status and not the symptom measure that drove the difference in the composite primary end point. These findings suggest that the validity of change in weight as a surrogate remains uncertain. Recent data from studies with implantable hemodynamic monitors have suggested that the relationship between hemodynamic congestion, total body volume retention (reflected as weight change), and symptoms is more complex than previously appreciated (43).

**Biomarkers.** The natriuretic peptides, which include B-type natriuretic peptide and its amino terminal fragment, have been established as important diagnostic tools in AHFS (44,45). Because of their association with heart failure disease severity and their wide clinical availability, natriuretic peptide levels are attractive as potential surrogate end points. Because a primary driver of natriuretic peptide levels is hemodynamic stress, many of the same concerns that apply to hemodynamic surrogates such as wedge pressure might also apply to natriuretic peptides. However,
The interaction between AHFS and the renal function.

In the SURVIVE (Survival of Patients with Acute Heart Failure in Need of Intravenous Inotropic Support) and the REVIVE-2 (Second Randomized Multicenter Evaluation of Intravenous Levosimendan Efficacy) programs, decreases in natriuretic peptide levels were assessed as important secondary end points (26,46). In both studies, levosimendan was associated with a substantial decrease in natriuretic peptide levels compared with placebo (in REVIVE-2) and dobutamine (in SURVIVE). In REVIVE-2, this favorable impact of levosimendan on reducing natriuretic peptide levels was associated with improvement in clinical symptoms and less worsening heart failure at 5 days but also corresponded with an observed increase in adverse events (hypotension and arrhythmias) and a trend toward increased long-term mortality. In the SURVIVE study, levosimendan did not result in improved short- or long-term clinical outcomes compared with dobutamine, despite a reduction in natriuretic peptide levels. At present, it remains unknown whether a change in biomarker levels might play a role as a secondary end point (26,46). In both studies, levosimendan was associated with a substantial decrease in natriuretic peptide levels compared with placebo (in REVIVE-2) and dobutamine (in SURVIVE). In REVIVE-2, this favorable impact of levosimendan on reducing natriuretic peptide levels was associated with improvement in clinical symptoms and less worsening heart failure at 5 days but also corresponded with an observed increase in adverse events (hypotension and arrhythmias) and a trend toward increased long-term mortality. In the SURVIVE study, levosimendan did not result in improved short- or long-term clinical outcomes compared with dobutamine, despite a reduction in natriuretic peptide levels. At present, it remains unknown whether a change in biomarker levels might play a role as a surrogate end point for dose selection or proof of concept in early phase studies in AHFS.

Renal function. The interaction between AHFS and the kidney has become a topic of substantially increased interest in recent years (47). Worsening renal function during hospital stay for AHFS, sometimes termed the “cardiorenal syndrome,” has been shown to be a powerful predictor of adverse outcomes in AHFS patients (48), and new therapies such as the adenosine antagonists have targeted preservation of renal function as a therapeutic goal in AHFS. Thus, although for many end points the distinction between a clinical end point and a surrogate is straightforward, measures of renal function seem to represent a grey zone between clinical end points, surrogate end points, and safety end points. In the ongoing rolofylline development program, worsening renal function (defined as a change in serum creatinine of ≥0.3 mg/dl) is a component of the composite primary end point (49). Although renal function could be considered a “safety end point,” a potentially useful distinction can be made between demonstrating safety (i.e., showing that renal function is not worsened by a novel therapy compared with control) and demonstrating efficacy (i.e., showing that renal function is preserved by the new therapy compared with placebo).

Safety End Points

Given the history of drug development in AHFS, the overall safety profile of new AHFS therapies has become an issue of significant concern. In particular when the focus of a therapy is short-term symptom relief, establishing that this does not occur at the expense of longer-term safety is critical. In this sense, safety end points are effectively noninferiority measures, requiring specific statistical approaches in order to establish noninferiority with a prespecified “equivalence boundary” (50). Evaluation of safety for new therapies should be guided by an understanding of the drug mechanism as well as by signals from earlier clinical study (e.g., renal dysfunction with nesiritide, ischemia with inotropic agents). 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). Additionally, planning of phase III studies should include formal assessment of the upper boundary of risk (either relative or absolute) that can be excluded by the planned sample size. Although appropriately planned phase III safety evaluation is invaluable, phase IV safety surveillance remains critical, because even with large sample sizes rare but clinically relevant increases in severe adverse events might not be fully excluded. For example, despite a sample size of 7,000 patients, the ongoing ASCEND-HF study is powered to detect (with 95% confidence and 90% power) a hazard ratio of 1.47 for 30-day mortality (assuming a baseline 30-day mortality rate of 4%). Ultimately, the degree of risk that must be “ruled out” 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 might be held to a higher standard of safety than 1 with more substantive clinical benefits).

Multiple and Composite End Points: Putting it Together

There is no single end point that accurately captures the totality of the patient experience with AHFS. Thus, substantial interest exists for combining end points 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 primary end points. Using more than 1 primary end point requires appropriate adjustment for multiple statistical comparisons. Typically, this takes the form of allocating the alpha (and, the potential for type I error) among the various end points. For example, in the ongoing ASCEND-HF study of nesiritide, the alpha is allocated between the co-primary end points of 1) death or heart failure hospital stay at 30 days (alpha = 0.045); and 2) dyspnea assessment at 6 and 24 h (alpha = 0.005) (23).

Distinct from the use of multiple primary end points, composite end points attempt to combine the various aspects of the AHFS syndrome into a single integrated measure. In their simplest form, a composite end point simply combines 2 separate “events” into 1 category (e.g., cardiovascular death or heart failure repeat hospital stay). Statistically, the use of composites increases the total event rate and therefore might...
increase statistical power. Importantly, however, composite end points only increase statistical power if the intervention has an effect on multiple aspects of the composite. The inclusion of factors in the composite end point that are not impacted by the intervention might actually “dilute” the observed treatment effect and decrease the overall statistical power (52).

A variety of complex composites have been proposed that try to incorporate many aspects of the patient experience into a single end point. In the REVIVE-2 study of levosimendan, the primary end point was the classification of the patient as “better, the same, or worse” with a pre-specified definition that included symptoms, worsening heart failure events, and death over a 5-day period (26) (Table 2). Patients with improved symptoms (moderate or marked improvement at 6 h, 24 h, and 5 days) and no worsening were classified as improved, whereas patients dying, experiencing worsening heart failure requiring rescue therapy, or experiencing worsening symptoms were classified as worse. Patients classified as neither better nor worse were classified as unchanged. Notably, the REVIVE trials demonstrated improvement in this end point (p = 0.015), but levosimendan was not approved in the U.S. due to concerns about hypotension and a possible trend toward late mortality in the levosimendan group. The ongoing phase III PROTECT I and II (Study of the Selective A1 Adenosine Receptor Antagonist KW-3902 [rolofylline] for Patients Hospitalized With Acute HF and Volume Overload to Assess Treatment Effect on Congestion and Renal Function) trial uses a similar “trichotomous” end point that incorporates symptom relief, worsening heart failure and/or renal function, and mortality (28).

Another alternative to these composite end points is hierarchical end points based on ranking of events, sometimes termed the “global rank approach.” In this type of scheme, all patients participating in a clinical trial are ranked on the basis of a pre-specified hierarchy of events. For example, time to death would be ranked at the bottom, then repeat hospital stay for heart failure, and so forth. Patients not experiencing any of these hard events could be ranked on the basis of a continuous measure (e.g., change in a symptom measure or a surrogate such as natriuretic peptides). The primary analysis in this type of analysis is the nonparametric comparison of the ranks between those patients in the intervention group versus those in the control group. One advantage of this type of end point is that it “weighs” the components of the clinical experience in a way that is generally congruent with clinical judgment and patient-perceived worth (mortality most important, nonfatal events next, symptoms next, and so on). The advantages and disadvantages of this type of global rank end point for trials of mechanical cardiac support devices have been reviewed in detail (53).

### Regulatory Issues

Regulatory requirements for approval of new drugs for AHFS have evolved rapidly over the past decade in response to many of the AHFS trial failures outlined in the preceding text. Currently, harmonization of regulatory agencies is lacking, often resulting in AHFS studies with multiple primary end points designed to meet divergent regulatory requirements. For example, the EMEA has neither embraced repeat hospital stay as an important end point in AHFS nor accepted composite end points. Therefore, approval requires either improvement in a symptom-based end point or mortality. By limiting the end points of a clinical trial to these opposite ends of the spectrum, the default has often been a symptom-based end point. In contrast, the FDA has been willing to accept composite end points and has encouraged some form of safety evaluation. This lack of regulatory harmonization has encouraged the use of co-primary end points and overly complex trial designs in order to meet divergent regulatory requirements (22,23).

### Key Considerations for Future End Point Design

The historical perspective gained by reviewing recent AHFS trials helps define general principles to guide future AHFS end point selection (Table 3). First and foremost, phase III trials of AHFS therapies must focus on measures of clinical importance assessed over a reasonable duration (54). Post-discharge clinical events such as mortality and major non-fatal events (such as repeat hospital stay) are of unequivocal clinical importance, and reducing these events is a major unmet clinical need in AHFS patients. Thus, treatments designed to improve these outcomes should be a primary goal of therapeutic development in AHFS. As noted earlier, whether traditional time-to-event analysis or alternate strategies for capturing these events (e.g., days alive and out of hospital) are preferable might differ depending on the nature of the intervention being tested, but the general principle remains the same. Substantial clinical changes during the index hospital stay (such as worsening heart failure and/or the need for rescue therapy) also capture an important element of the clinical course and are valid for measuring the clinical benefits of therapy, so long as they are accompanied by longer-term evidence of safety. Although

<table>
<thead>
<tr>
<th>Table 3</th>
<th>Key Considerations for an Ideal Acute Heart Failure Syndromes End Point</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Measure something of clinical value (i.e., relates to patients feeling better and/or living longer or saves resources)</td>
</tr>
<tr>
<td>2.</td>
<td>Assess over a reasonable time period (i.e., more than 24 h)</td>
</tr>
<tr>
<td>3.</td>
<td>Validate measurement tools (e.g., symptom scales)</td>
</tr>
<tr>
<td>4.</td>
<td>Account for the relative value of each component in composites (i.e., ranking methods)</td>
</tr>
<tr>
<td>5.</td>
<td>Standardize key end points between trials</td>
</tr>
<tr>
<td>6.</td>
<td>Harmonize regulatory agency expectations (i.e., Food and Drug Administration and European Medicines Agency)</td>
</tr>
</tbody>
</table>
symptom relief remains an important goal of therapy, symptomatic improvement (in the absence of other clinical benefits) must be rapid, substantial, and sustained beyond the initial hours of treatment to be considered a significant therapeutic advance over usual care. Dyspnea relief is best considered in the context of other short- and intermediate-term benefits, suggesting that composite end points such as those from REVIVE-2 and the PROTECT program are superior to isolated dyspnea measures in demonstrating efficacy (26,49). Finally, safety remains a critical challenge for evaluating new AHFS therapies. Given the size of experience needed to provide definitive evidence of safety, regulatory agencies should consider providing initial approval but with specific requirements for post-marketing safety surveillance with valid safety end points and appropriate studies to judge the relative risks of therapeutic versus usual care.

Moving forward, standardization and validation of end point measures is critical. The call for a common dyspnea end point that is objective and then validated is a beginning, but such efforts should extend to other AHFS end point measures. Greater agreement and harmonization between clinical trialists, industry sponsors, and regulatory agencies will also be required in order to evaluate new therapies in the most efficient way possible.

Conclusions

The field of AHFS trial design continues to evolve. Reliance on short-term hemodynamic benefits or other surrogate end points is clearly unacceptable as a means of establishing definitive clinical efficacy. Moving forward, phase III clinical trials in AHFS will need to focus on measures that clearly reflect clinical efficacy (mortality, hospital stay, worsening heart failure, and/or clinically meaningful and durable relief of symptoms) as well as provide definitive evidence of longer-term safety. This will almost certainly require larger trials than those previously used to evaluate new AHFS therapies (thousands of patients rather than hundreds of patients). Studies such as ASCEND-HF can establish a precedent for doing trials adequately powered to provide definitive evidence regarding both efficacy and safety and provide a stage for doing trials adequately powered to provide definitive evidence of both efficacy and safety and provide a stage for doing trials adequately powered to provide definitive evidence of both efficacy and safety and provide a stage for doing trials adequately powered to provide definitive evidence of both efficacy and safety and provide a stage for doing trials adequately powered to provide definitive evidence of both efficacy and safety and provide a stage for doing trials adequately powered to provide definitive evidence of both efficacy and safety and provide a stage for doing trials adequately powered to provide definitive evidence of both efficacy and safety and provide a stage for doing trials adequately powered to provide definitive evidence of both efficacy and safety and provide a stage for doing trials adequately powered to provide definitive evidence of both efficacy and safety and provide a stage for doing trials adequately powered to provide definitive evidence of both efficacy and safety and provide a stage for doing trials adequately powered to provide definitive evidence of both efficacy and safety and provide a stage for doing trials adequately powered to provide definitive evidence of both efficacy and safety and provide a stage for doing trials adequately powered to provide definitive evidence of both efficacy and safety and provide a stage for doing trials adequately powered to provide definitive evidence of both efficacy and safety and provide a stage for doing trials adequately powered to provide definitive evidence of both efficacy and safety and provide a stage for doing trials adequately powered to provide definitive evidence of both efficacy and safety and provide a stage for doing trials adequately powered to provide definitive evidence of both efficacy and safety and provide a stage for doing trials adequately powered to provide definitive evidence of both efficacy and safety and provide a stage for doing trials adequately powered to provide definitive evidence of both efficacy and safety and provide a stage for doing trials adequately powered to provide definitive evidence of both efficacy and safety and provide a stage for doing trials adequately powered to provide definitive evidence of both ef

REFERENCES


PROTECT 1 and 2: A Study of the Selective A1 Adenosine Receptor Antagonist KW-3902 for Patients Hospitalized With Acute HF and Volume Overload to Assess Treatment Effect on Congestion and Renal Function. Available at: http://clinicaltrials.gov/ct2/show/NCT00354458?intr
