Development of Therapeutics for Heart Failure

When to Stop a Clinical Trial Early for Benefit: Lessons Learned and Future Approaches

Faiez Zannad, MD, PhD; Wendy Gattis Stough, PharmD; John J.V. McMurray, MD; Willem J. Remme, MD; Bertram Pitt, MD; Jeffrey S. Borer, MD; Nancy L. Geller, PhD; Stuart J. Pocock, PhD

Early stopping of a clinical trial for evidence of benefit has been widely debated in the medical literature.1–13 This practice has important implications from many viewpoints: clinicians who practice evidence-based medicine; future patients to whom the results of research studies apply; patients who voluntarily agree to participate in clinical trials; and scientists, investigators, and regulators who strive to balance conducting scientifically rigorous studies with disseminating data that support therapeutic advances as quickly as is reasonable.

During the 7th Global Cardiovascular Clinical Trialists Forum held in Paris, France, in December 2010, cardiovascular clinical trialists, biostatisticians, National Institutes of Health scientists, regulators, and pharmaceutical industry scientists met to discuss current issues related to cardiovascular clinical trials, including the topic of stopping a clinical trial early for benefit. The Eplerenone in Mild Patients Hospitalization and Survival Study in Heart Failure (EMPHASIS-HF) is a recent trial that was stopped early for benefit14 and was used as a stimulus for discussion. This report summarizes the results of the group’s discussion on the scientific, statistical, and practical issues regarding the topic of stopping a clinical trial early for benefit.

The EMPHASIS-HF Experience
EMPHASIS-HF was a randomized, double-blind, clinical trial of eplerenone compared with placebo, in addition to maximally tolerated doses of an angiotensin-converting enzyme inhibitor, angiotensin receptor blocker, and β-blocker (unless contraindicated) in patients with mild (New York Heart Association class II) heart failure symptoms and left ventricular ejection fraction ≤30% (≤35% was allowed for patients with QRS duration >130 ms).14 The primary end point was death from cardiovascular causes or hospitalization for heart failure. This trial was of particular clinical relevance because it was the first trial of a mineralocorticoid receptor antagonist in heart failure patients with mild symptoms.

EMPHASIS-HF was monitored by an independent data monitoring committee (DMC) with 2 prespecified interim analyses and a stopping guideline of $P<0.001$ for benefit on the primary end point. The study initially planned to enroll 2584 patients when recruitment began in March 2006. The sample size was increased to 3100 patients in June 2009 because the blinded overall event rate was lower than anticipated.

In May 2010, after the second interim analysis, the DMC reported to the executive committee chairs that the prespecified stopping boundary had been crossed, with benefit favoring eplerenone as regards the primary composite end point of cardiovascular death or heart failure hospitalization, analyzed as time to first event. Based on this information, the full executive committee decided to stop the trial in May 2010 after 2737 patients had been enrolled. The available interim results at that time included a total of 559 patients with a reported primary event (231 eplerenone and 328 placebo, log-rank $P<0.00001$). Ninety percent of these events had been adjudicated by the clinical end point committee (CEC). The final published results were consistent with the findings from the interim analyses. The eplerenone group had 249 patients with a primary event compared with 356 in the placebo group ($P<0.00001$); 171 patients died in the eplerenone arm compared with 213 in the placebo arm ($P=0.008$).

The decision to stop the EMPHASIS-HF trial was made on the basis of several principles. First, the decision was consistent with the prespecified stopping guidelines in the DMC charter. Second, the level of statistical significance observed on the interim analysis minimized concerns that the finding might reverse or reflect the play of chance. Third, the findings were consistent when the components of the primary end point were evaluated individually. Finally, the results were

Received July 19, 2011; accepted December 22, 2011.
From INSERM, Centre d’Investigation Clinique 9501 and Unité 961, Centre Hospitalier Universitaire, and the Department of Cardiology, Nancy University, Nancy, France (F.Z.); Campbell University College of Pharmacy and Health Sciences, Buies Creek, NC, and Duke University Medical Center, Durham, NC (W.G.S.); Western Infirmary and the British Heart Foundation, Glasgow Cardiovascular Research Centre, University of Glasgow, Glasgow, United Kingdom (J.J.V.M.); Sticares Cardiovascular Research Institute, Rhoon, The Netherlands (W.J.R.); University of Michigan School of Medicine, Ann Arbor, MI (B.P.); State University of New York Downstate Medical Center, Brooklyn and New York, NY (J.B.); National Heart, Lung, and Blood Institute, Bethesda, MD (N.G.); and the Department of Medical Statistics, London School of Hygiene and Tropical Medicine, London, United Kingdom (S.J.P.).

Correspondence to Faiez Zannad, MD, PhD, FESC, CIC INSERM CHU, Hôpital Jeanne d’ Arc, 54200 Toul, Nancy, France. E-mail f.zannad@chu-nancy.fr

(Circ Heart Fail. 2012;5:294-302.)

© 2012 American Heart Association, Inc.

Circ Heart Fail is available at http://circheartfailure.ahajournals.org

DOI: 10.1161/CIRCHEARTFAILURE.111.965707
consistent with previous trials of mineralocorticoid receptor antagonists.\textsuperscript{15,16}

One concern about stopping early raised during the executive committee and DMC discussions was the possibility of missing a beneficial effect of eplerenone on survival. At the time of the second interim analysis, some evidence of improved survival was observed, but the strength of evidence was borderline ($P=0.044$). Consideration was given to continuing the trial until the survival benefit was more definitive because of the possibility that the evidence of survival benefit might not persist once all data were collected and analyzed (eg, the end result could be $P>0.05$). The rationale in support of not stopping was based on the opinion that if eplerenone indeed improved survival as suggested by the interim data, then it would be important to enhance the degree of confidence around the benefit. The argument to continue the trial to fully evaluate the effect of eplerenone on mortality was compelling because no pharmacological agent had been shown to improve survival in patients with heart failure and mild symptoms since the emergence of the \textbeta-adrenergic blockade data in the late 1990s. Nevertheless, the executive committee ultimately decided to stop the trial early. Although mortality reduction was clearly an important matter, it was a secondary end point and was not the end point on which the prespecified stopping guidelines were based. Importantly, the executive committee believed that, given the marked effect on the primary end point that was clearly established beyond reasonable doubt, there was no longer clinical equipoise, raising ethical issues for continuing the trial, particularly for those patients randomly assigned to placebo. Additionally, the integrity of the trial could have been compromised had the trial continued because unblinded results were known to the executive committee chairs.

### Challenges and Implications for Stopping a Trial Early

**End Point Considerations**

Determining the most appropriate choice of end points for the primary efficacy outcome and stopping guidelines is a critical decision point for executive committees and DMC members. Clinical relevance is, of course, a primary driver of end point selection. However, other operational factors such as achievable sample sizes, expected event rates, and intended duration of a trial also play a role. Thus, executive committees may choose to assign a composite for the primary end point and evaluate all-cause mortality as a secondary end point. It may be reasonable for DMCs and executive committees to consider using mortality to define the stopping boundary in such cases rather than or in addition to the primary composite end point. This may be especially important when evaluating mortality is of particular clinical interest, because there is a risk of missing an important effect on secondary end points (such as mortality) when trials are stopped early for benefit.

Although a composite end point may be appropriate as a primary clinical end point, it may be less desirable as the only end point for predefined stopping boundaries for several reasons. First, endpoints that require subjective decision-making (revascularization or hospitalization) are sometimes difficult to interpret and probably should not be used to stop a trial early for benefit. Second, it is possible for an overall composite end point to demonstrate a favorable effect but the effect may be neutral or negative in one of the individual components. For a trial to stop early for benefit, the observed effect on the composite end point ideally should be consistent in the individual components. In the Heart Outcomes Prevention Evaluation (HOPE) trial, the primary composite end point was positive early in the trial, but the DMC waited to recommend stopping the study until all components of the primary were positive.\textsuperscript{17}

The Anglo-Scandinavian Cardiac Outcomes Trial–Blood Pressure Lowering Arm (ASCOT-BPLA) is another interesting case example of a trial stopped early for benefit.\textsuperscript{18} The study was designed to evaluate the superiority of an amlodipine-based antihypertensive regimen as compared with an atenolol-based regimen on the primary end point of nonfatal myocardial infarction and fatal coronary heart disease. The DMC recommended stopping the trial early after interim analyses indicated that patients randomly assigned to the amlodipine-based regimen had a lower incidence of fatal and nonfatal stroke (230 versus 390, $P<0.0001$). However, this was not part of the primary end point (myocardial infarction and fatal coronary heart disease, 315 versus 354, $P=0.14$). After much debate, the trial’s executive committee decided the trial should continue. One year later, the DMC again recommended stopping because there were significantly fewer deaths on the amlodipine-based regimen. The executive committee stopped the trial at that time despite the fact that the primary end point was still not significantly different between treatment groups. The final published results based on 19 257 patients with a median 5.5 years of follow-up were as follows for amlodipine versus atenolol: stroke, 327 versus 422, $P=0.0003$; myocardial infarction and fatal coronary heart disease, 429 versus 474, $P=0.11$ (the primary end point); cardiovascular death, 263 versus 342, $P=0.001$; and all-cause death, 738 versus 820, $P=0.02$.\textsuperscript{18} In this example, the primary end point included fatal coronary heart disease, an end point requiring adjudication of a specific cause of death. As is well known to clinical investigators, it is often very difficult to provide unambiguous determinations of causes of death. Thus, the recommendation of the DMC and the decision of the executive committee might have been influenced by the less ambiguous results relating to stroke and all-cause death.

A consistency of effect may be desirable for some combined end points and their individual components, but it may not be necessary for others. In a trial that used a combined end point of cardiovascular mortality and heart failure hospitalization, an effect on heart failure hospitalization without a strong trend on cardiovascular mortality would not be a compelling reason to stop the trial early for benefit, due to the subjectivity of the end point and its variability among patients and regions of the world. However, in a trial that used a combined end point of cardiovascular mortality or stroke, a strong effect on stroke might be compelling without a corresponding effect on death from other cardiovascular causes. DMC recommendations should not only depend on the primary composite end point, but it should also include careful consideration of the direction, magnitude, and strength of benefit in the
individual components (and all-cause mortality), as well as the potential for knowledge to be gained (or lost) if the trial is stopped early.

**Effect of Stopping Early on Knowledge of Precision and Magnitude of Effect**

Estimating the magnitude of effect of a new treatment, and the precision around that effect, is an important contribution of clinical research toward the advancement of patient care. Understanding these estimates is critical for number needed to treat analyses, guidelines development, and cost-effectiveness research. The ability to determine the true magnitude and precision of the estimate of the treatment effect may be lessened when trials are stopped early for benefit.

Bassler et al. demonstrate how effects may be overestimated when trials are stopped early. There is a lack of consensus among clinical trial experts about how statistical techniques and/or prespecified stopping guidelines can overcome this bias. Ensuring that stopping guidelines require very strong evidence (proof beyond a reasonable doubt) and allow for accrual of an adequate number of events are important considerations.

The Candesartan in Heart Failure Assessment of Reduction in Mortality and morbidity (CHARM) trial exemplifies how results can evolve as the evidence accumulates over time. The CHARM DMC’s guiding principle was that early termination would only be recommended when the evidence provided proof beyond a reasonable doubt that the results would change clinical practice. Thus, the predefined statistical stopping guidelines were based on all-cause mortality, with a required probability value of <0.001. The CHARM DMC also required a more stringent threshold (P<0.0001) for stopping within 18 months of the first patient randomly assigned in the trial, when the numbers of events were expected to be small. In retrospect, this provision was quite important. By the second interim analysis, a highly significant difference in mortality was observed in favor of candesartan, based on 76 deaths in the candesartan arm and 123 in the placebo arm (P=0.0007). Thus, the guideline for stopping was not reached since it was within the 18-month window. By the 3rd interim look, the probability value was 0.0002, which still did not reach the stopping boundary of <0.0001. Although the stopping boundary was reached at the 4th interim analysis, the DMC recommended that the study continue on the basis of several factors, one of which was the attenuation of the hazard ratio for mortality as more events had accrued.9 When the trial was completed, the treatment difference in mortality between the candesartan and placebo arms was of borderline significance. In 7599 patients followed for a median of 3.1 years, there were 886 versus 945 deaths (hazard ratio [HR], 0.91; 95% confidence interval [CI], 0.83–1.00; P=0.055). There was a highly significant difference in the composite of cardiovascular death and hospitalization for heart failure (HR, 0.84; 95% CI, 0.77–0.91; P<0.0001).

The CHARM data monitoring experience exemplifies how the magnitude of benefit and the estimate of treatment effect become more precise (and may sometimes attenuate) as additional events accrue. Careful consideration should be given to choosing time points for interim efficacy analyses because treatment effects tend to be exaggerated early in any trial. This phenomenon has been described as “regression to the truth.” The challenge for DMCs is to avoid stopping a trial after short follow-up, when the overestimation of effect is most likely to occur. Establishing a minimum follow-up period as part of the stopping guidelines is one approach to minimize this risk, and it is a strategy that is increasingly requested by regulatory agencies. DMCs and executive committees must weigh the importance of obtaining a precise result against the potential risks of allowing research subjects to continue receiving placebo. The former may be most important for a new class of therapy, or for a therapy with which an effect on a major clinically relevant end point has never been shown, or where safety may still be in question. In the Cardiac Insufficiency Bisoprolol Study II (CIBIS II), the DMC recommended the study be stopped after the first interim analysis revealed a beneficial effect for bisoprolol on all-cause mortality (P<0.001). However, the steering committee opted to continue the trial because of the uncertainty surrounding the safety and efficacy of β-blocker use in heart failure at that time. The study was subsequently stopped early after the 2nd interim analysis revealed similar findings.

Concerns about precision of results and magnitude of benefit have led to the criticism of some trials that were stopped early for benefit. The Justification for the Use of Statins in Primary Prevention: an Intervention Trial Evaluating Rosuvastatin (JUPITER) was stopped after the second interim analysis with 328 events and a median follow-up of 1.9 years. The study reported a 44% reduction in the primary end point of myocardial infarction, stroke, arterial revascularization, hospitalization for unstable angina, or death from cardiovascular causes for rosuvastatin as compared with placebo.20 Some have argued that stopping the trial early introduced bias and resulted in an unexpectedly large and rapid treatment effect, factors that inhibit the impact of the data on clinical practice. Another criticism is that “soft” end points were used as the basis to stop the trial. The study investigators point to the stringent DMC procedures, which required that any stopping be based on “statistically extreme findings” that supported “proof beyond reasonable doubt” in accordance with conservative stopping guidelines.22 The investigators noted that the DMC continued the trial for an additional period of 6 months to further evaluate the certainty of the magnitude of benefit and until all components of the end point and total mortality were also reduced significantly with very small probability values (myocardial infarction, stroke, and cardiovascular death: relative risk [RR], 0.53; 95% CI, 0.49–0.69; P<0.000001; revascularization or hospitalization for unstable angina: RR, 0.53; 95% CI, 0.40–0.70; P<0.000001; all-cause mortality: RR, 0.80; 95% CI, 0.67–0.97; P=0.02). The final probability value for the primary composite end point was <0.00000001.

How many events and what strength of evidence are enough to achieve adequate statistical certainty that stopping a trial early for benefit is appropriate? Numerous examples can be offered in which smaller studies have suggested large mortality benefits with significant probability values, but when larger studies were conducted, the mortality effect was
either neutral or in some cases harmful.24–26 Bassler et al1 rather arbitrarily suggest that the threshold should be 500 events. Agreement among experts has certainly not been reached that this is an appropriate number because the exact number depends on the trial’s specific context and previous experience with the drug.

DMCs must not only consider statistical stopping guidelines when making decisions regarding the early termination of a trial for benefit, but they also must consider the uptake and acceptance of the result by the clinical community. If the data are relatively insubstantial, clinicians are skeptical of the results, and the trial may be viewed as an unnecessary loss of time, resources and, potentially, lives. On the other hand, the desire to achieve scientific certainty must be balanced against the ethical need to provide the best care to study participants and to avoid delaying the wider awareness of therapeutic advances. Table 19,14,15,19,27–39 provides a selected summary of some cardiovascular trials stopped early for benefit and the implications of stopping.

Totality of the Evidence
The totality of evidence is another important consideration for DMCs when deciding to recommend stopping early.40–42 Primary, secondary, and safety end points should be considered, as should the consistency of the effect across multiple (preferably prespecified) subgroups. Data from previous relevant trials can be used to assess the external validity of the results. It may also be available from concurrent trials with the same drug or drug class, but sharing interim data to assess benefit is generally not advisable. The appropriateness of sharing interim results among DMCs was discussed at a previous Cardiovascular Clinical Trialists workshop.43 Recommending early termination for a trial of a novel, first-in-class agent should be done cautiously because the ability to externally replicate the finding will be limited or may not exist. In these cases, emphasis should be placed on ensuring an adequate quantity and strength of evidence have accrued and that the length of follow-up is sufficient, to avoid the potential of stopping the trial on a spurious “random high.”10,13 Conversely, for new members of an existing drug class in which previous positive trials or meta-analyses exist, the broader totality of evidence can be reviewed. This practice will help reduce the risk of any exaggerated claim of therapeutic efficacy.

The Ivabradine and Outcomes in Chronic Heart Failure (SHIFT) trial was not stopped early. Although a significant effect was observed at the second interim efficacy analysis on the end point of cardiovascular death and heart failure hospitalization (HR, 0.77; \( P<0.0001 \), thus crossing the prespecified significance level of \( P=0.001 \)) as well as all-cause mortality (HR, 0.77; \( P=0.0014 \)), the trial was continued, in part because a previous study (in a different population) was negative.44 In addition, ivabradine was a new agent in a drug class not previously studied in heart failure; thus, on balance, the totality of evidence was deemed by the DMC not to be sufficiently convincing to justify stopping the trial early. It should be noted that the effect size of the final results were less than that observed in the interim analysis (cardiovascular death and heart failure hospitalization: HR, 0.82; 95% CI, 0.75–0.90; \( P<0.0001 \); all-cause mortality: HR, 0.90; 95% CI, 0.80–1.02; \( P=0.092 \)).

Responsibility to Subjects: Treating Participants After Stopping a Trial Early
Stopping a trial before its planned duration is associated with operational challenges. When a trial is stopped early for benefit, investigators have an ethical responsibility to offer the more effective treatment to all study participants. This process is less complicated for trials of agents with existing regulatory approval. However, the introduction of open-label therapy must be conducted under the review and approval of the institutional review board for investigational drugs and for approved drugs if the labeled indication differs from the disease state under study. The cost of therapy may also be a limitation for some patients because insurance reimbursement policies will lag behind the emerging clinical trial evidence.

In EMPHASIS-HF, an eplerenone open-label extension phase was not included in the original study design. Because an amendment had to be reviewed by local institutional review boards, a lengthy interval elapsed between when the study was stopped and when open-label eplerenone was available to some patients. In an effort to reduce such delays, executive committees may consider including plans for transitioning patients from double-blind to open-label drug in the original protocol, in the event the study is stopped early for benefit. In addition, local institutional review or ethics boards should have processes in place that allow them to expedite the review of open-label extension studies when evidence of benefit has been generated that is sufficient to stop a trial early.

Rapid collection and adjudication of remaining events is another challenge facing trials that stop early. If a trial is highly positive, particularly for critical end points such as mortality, then it is desirable to complete final collection, analysis, and dissemination of the data as quickly as possible. During the trial planning phase, the study operations team should develop processes for data collection, query resolution, and event adjudication that take place on an ongoing basis. The goal of these processes should be to receive data and resolve data queries quickly and to have minimal delay between the reporting of events and their adjudication. In large clinical trials, in which multiple organizations are often involved, these processes consume substantial amounts of time, and backlogs can and do occur. When a trial is stopped unexpectedly, these backlogs become even more pronounced. Methods to streamline these functions while maintaining quality should be preplanned so that study close-out can be as short as possible if a trial is stopped early.

Proposed Approaches for Future Trials
There is no substitute for a DMC charter that specifies DMC responsibilities, guidelines for early termination (for benefit or for harm), and methods for communication of interim results to the executive committee (Table 2).43,45 An experienced and well-chosen DMC is also needed that can make wise judgments when faced with complex issues and evidence pertinent to a potential recommendation to stop (or not stop) early. This requires that the DMC be able to evaluate the
Table 1. Selected Examples of Heart Failure Trials That Did and Did Not Stop Early for Benefit

<table>
<thead>
<tr>
<th>Trial</th>
<th>Stopping Procedures</th>
<th>End Point for Stopping</th>
<th>Length of Follow-Up</th>
<th>Stopped Early, Yes or No</th>
<th>Final Results</th>
<th>Clinical Implications (Acceptance/Uptake)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CONSENSUS37</td>
<td>No formal stopping rule in place</td>
<td>DMC reviewed data every 3 mo for mortality</td>
<td>Actual: 188 d (mean); planned: 12 mo</td>
<td>Yes</td>
<td>n = 253&lt;br&gt; All-cause mortality: 50 (39%) enalapril vs 68 (54%) placebo, P = 0.003</td>
<td>Cornerstone of HF management, guideline-recommended therapy</td>
</tr>
<tr>
<td>GESICA38</td>
<td>DMC reviewed after one-third and two-thirds of planned enrollment</td>
<td>All-cause mortality</td>
<td>Actual: 13 mo (mean); planned: 2 y</td>
<td>Yes</td>
<td>n = 516&lt;br&gt; Mortality: 87 amiodarone vs 106 control; RR, 28%; 95% CI, 4–45%, P = 0.024</td>
<td>Stopped on random high, “regression to truth” with CHF STAT39</td>
</tr>
<tr>
<td>CHF STAT30</td>
<td>Stopping procedures not described in results report</td>
<td>Not specified</td>
<td>Actual: 45 mo (median); Planned: Minimum 1 y</td>
<td>No</td>
<td>n = 674&lt;br&gt; Mortality: 131 (39%) amiodarone vs 143 (42%) placebo, P = 0.6</td>
<td>Follow-up &gt;=3 x longer, more patients, more events than GESICA</td>
</tr>
<tr>
<td>US Carvedilol Trials32</td>
<td>DMC met periodically to review blinded results, no formal rules for stopping</td>
<td>All-cause mortality for the stratified trial program (4 individual trials with nonfatal primary end points)</td>
<td>Actual: 6.5 mo (median); Planned: 12 mo</td>
<td>Yes</td>
<td>n = 1094&lt;br&gt; Mortality: 22 (3.2%) carvedilol vs 31 (7.8%) placebo; P &lt; 0.001</td>
<td>At time of publication, much debate in the medical community regarding the robustness of these findings; small numbers of deaths, short follow-up were some of the criticisms</td>
</tr>
<tr>
<td>CIBIS-II33</td>
<td>Peto rule P &lt; 0.001 for all-cause mortality</td>
<td>All-cause mortality</td>
<td>Actual: 1.3 y (mean); Planned: 2 y</td>
<td>Yes</td>
<td>n = 2647&lt;br&gt; Mortality: 156 (11.8%) bisoprolol vs 228 (17.3%) placebo; RR, 0.66; 95% CI, 0.54–0.8, P = 0.0001</td>
<td>Results consistent with totality of evidence</td>
</tr>
<tr>
<td>MERIT-HF34</td>
<td>Asymmetric group-sequential procedure, with cumulative probability of stopping early of 0.0036</td>
<td>All-cause mortality</td>
<td>Actual: 1 y (mean); Planned: 2.4 y</td>
<td>Yes</td>
<td>n = 3991&lt;br&gt; Mortality: 145 (7.2%) metoprolol XL vs 217 (11%) placebo; RR, 0.66; 95% CI, 0.53–0.81, P = 0.0062</td>
<td>Results consistent with totality of evidence</td>
</tr>
<tr>
<td>COPERNICUS32</td>
<td>Truncated O'Brien-Fleming type boundary computed with the lan-DeMets procedure</td>
<td>All-cause mortality</td>
<td>Actual: 10.4 mo (mean); Planned: event driven to 900 deaths</td>
<td>Yes</td>
<td>n = 2289&lt;br&gt; Mortality: 130 (11.4%) carvedilol vs 190 (18.5%) placebo; RR, 35%; 95% CI, 19–48%, P = 0.0014</td>
<td>Wide acceptance, even in patients not reflected among study population</td>
</tr>
<tr>
<td>RALES15</td>
<td>Group sequential monitoring plan with Lan-DeMets stopping boundary and an O'Brien Fleming spending function</td>
<td>All-cause mortality</td>
<td>Actual: 24 mo (mean); Planned: 3 y</td>
<td>Yes</td>
<td>n = 1663&lt;br&gt; Mortality: 284 (35%) spironolactone vs 386 (46%) placebo; RR, 0.7; 95% CI, 0.6–0.82, P &lt; 0.001</td>
<td></td>
</tr>
<tr>
<td>A-HeFT33</td>
<td>Lan DeMets sequential boundaries</td>
<td>All-cause mortality (different from primary composite end point of all-cause death, HF hospitalization, and change in quality of life)</td>
<td>Actual: 10 mo (mean); Planned: 18 mo</td>
<td>Yes</td>
<td>n = 1050&lt;br&gt; Mortality: 32 (6.3%) Bidil vs 54 (10.2%) placebo; HR, 0.57, P = 0.01</td>
<td>Accepted and incorporated as guideline-recommended therapy, although uptake has been suboptimal</td>
</tr>
<tr>
<td>CHARM32,34</td>
<td>Haybittle-Peto rule, requiring 2-sided test stratified by trial, for interim analyses within 18 mo of randomization, P &lt; 0.001 was required</td>
<td>All-cause mortality (different from primary end point of overall trial, which was CV death or HF hospitalization)</td>
<td>Actual: 37.7 mo (median); Planned: 2 y (minimum)</td>
<td>No</td>
<td>n = 7599&lt;br&gt; Primary (CV death or HF hospitalization); adjusted HR, 0.62; 95% CI, 0.77–0.91, P &lt; 0.0001</td>
<td>Guideline-recommended therapy in ACE, inhibitor-intolerant patients or as add-on therapy</td>
</tr>
</tbody>
</table>

(Continued)
data at each interim analysis as well as trends with the accruing data. Whether or not data from ongoing external trials can be shared or considered among DMCs has been discussed previously, but it is generally not advised for purposes of efficacy evaluations.43 With regard to the decision to stop early, the DMC charter should specify the roles and responsibilities of the DMC and whether they are an advisory or decision-making body. This issue becomes relevant if the executive committee chooses not to accept the DMC recommendation to stop (or not to stop) early. This possibility should be considered in the charter. Procedures should be in place, if possible, that would allow the executive committee to maintain clinical equipoise, because once the interim findings are made known to the executive committee, continuation of the trial could negatively affect the study’s integrity. Of similar importance is the issue of whether or not

Table 1. Continued

<table>
<thead>
<tr>
<th>Trial</th>
<th>Stopping Procedures</th>
<th>End Point for Stopping</th>
<th>Length of Follow-Up</th>
<th>Stopped Early, Yes or No</th>
<th>Final Results</th>
<th>Clinical Implications (Acceptance/Uptake)</th>
</tr>
</thead>
<tbody>
<tr>
<td>EMPHASIS14</td>
<td>Adaptation of the Peto stopping criterion; interim analyses planned after 271 and 542 primary end point events accrued; termination could be recommended after 542 primary end points if 2-sided P&lt;0.001</td>
<td>CV death or HF hospitalization</td>
<td>Actual: 21 mo (median) Planned: 48 mo</td>
<td>Yes</td>
<td>n=2737</td>
<td>Primary (CV death or HF hospitalization): 249 (18.3%) eplerenone vs 356 (25.9%) placebo; adjusted HR, 0.63; 95% CI, 0.54–0.74, P&lt;0.001 All-cause mortality: 171 (12.5%) eplerenone vs 213 (15.5%) placebo, adjusted HR, 0.76; 95% CI, 0.62–0.93, P=0.008</td>
</tr>
<tr>
<td>SHIFT25</td>
<td>Peto procedure P=0.001 at each of 2 interim analyses</td>
<td>CV death or HF hospitalization</td>
<td>22.9 mo (median)</td>
<td>No</td>
<td>n=6505</td>
<td>CV death or HF hospitalization: 793 (24%) ivabradine vs 937 (29%) placebo; HR, 0.82; 95% CI, 0.75–0.90, P&lt;0.0001 All-cause mortality: 503 (16%) ivabradine vs 552 (17%) placebo; HR, 0.90; 95% CI, 0.80–1.02, P=0.092</td>
</tr>
</tbody>
</table>

Heart failure device trials

<table>
<thead>
<tr>
<th>Trial</th>
<th>Stopping Procedures</th>
<th>End Point for Stopping</th>
<th>Length of Follow-Up</th>
<th>Stopped Early, Yes or No</th>
<th>Final Results</th>
<th>Clinical Implications (Acceptance/Uptake)</th>
</tr>
</thead>
<tbody>
<tr>
<td>MADIT26</td>
<td>Triangular sequential design with prespecified stopping boundaries</td>
<td>All-cause mortality</td>
<td>Actual: 27 mo Planned: 5 y</td>
<td>Yes</td>
<td>n=196</td>
<td>Mortality: 15/95 defibrillator vs 39/101 conventional treatment; HR, 0.46; 95% CI, 0.26–0.82, P=0.009</td>
</tr>
<tr>
<td>MADIT IF27</td>
<td>Triangular sequential design with prespecified stopping boundaries</td>
<td>All-cause mortality</td>
<td>Actual: 20 mo (mean) Planned: 2 y</td>
<td>Yes</td>
<td>n=1232</td>
<td>Mortality: 105 (14.2%) ICD vs 97 (19.8%) no ICD; HR, 0.69; 95% CI, 0.51–0.93, P=0.016</td>
</tr>
<tr>
<td>COMPANION38</td>
<td>O-Brien Fleming implemented by Lan and DeMets</td>
<td>All-cause death or all-cause hospitalization</td>
<td>Actual: 14.8–16.5 mo (median) Planned: Event-driven (1000 events)</td>
<td>Yes</td>
<td>n=1520</td>
<td>All-cause death or hospitalization was 68% pharm vs 56% CRT; HR, 0.81; 95% CI, 0.69–0.96, P=0.014, vs 56% CRT-D; HR, 0.8; 95% CI, 0.68–0.95, P=0.01 Mortality alone: pharm (77 deaths) vs CRT (131 deaths); HR, 0.76; 95% CI, 0.58–1.01; P=0.059, vs CRT-D (105 deaths); HR, 0.64; 95% CI, 0.48–0.86, P=0.003</td>
</tr>
<tr>
<td>MADIT-CRT39</td>
<td>Wang-Tsiatis group sequential design; 20 interim analyses with approximately 35 events each</td>
<td>All-cause death or nonfatal HF events</td>
<td>Actual: 2.4 y (mean) Planned: 2 y (minimum)</td>
<td>Yes</td>
<td>n=1820</td>
<td>Deaths or HF events: 187 (17.2%) CRT-ICD vs 185 (25.3%) ICD only; HR, 0.66; 95% CI, 0.52–0.84, P=0.001</td>
</tr>
</tbody>
</table>

DMC indicates data monitoring committee; HF, heart failure; RR, relative risk; CI, confidence interval; CHF, congestive heart failure; HR, hazard ratio; ACE, angiotensin-converting enzyme; CV, cardiovascular; ICD, implantable cardioverter-defibrillator; CRT, cardiac resynchronization therapy; NYHA, New York Heart Association.
Table 2. Considerations for Stopping a Trial Early

<table>
<thead>
<tr>
<th>Considerations (with respect to decisions to stop early)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Role of DMC as advisory or decision-making body and to whom it should report</td>
</tr>
<tr>
<td>Need for full agreement of DMC for early stopping recommendations. If full agreement not required, proportion of members that must be in agreement</td>
</tr>
<tr>
<td>Predefined statistical stopping boundaries</td>
</tr>
<tr>
<td>Clinical parameters that will be considered in decision</td>
</tr>
<tr>
<td>Number of events required</td>
</tr>
<tr>
<td>Length of follow-up required</td>
</tr>
<tr>
<td>Level of significance required</td>
</tr>
<tr>
<td>Totality of evidence</td>
</tr>
</tbody>
</table>

Protocol considerations

- Procedures to communicate with subjects and rapidly schedule close-out visits
- Procedures to rapidly communicate with ethics committees
- Plan to switch patients to active drug (if effective)
- Language in the original consent to cover the possibility of an open-label, active-drug extension phase, to minimize the lag time in transitioning subjects to effective therapy
- Procedures to rapidly collect outstanding data and resolve data queries
- Rapid collection of source documents needed to finalize event adjudication
- Process to rapidly address any pending adjudication assessments
- Process for expedited adjudication while maintaining high-quality standards
- Confidentiality plan that allows release of necessary information to treat patients appropriately without jeopardizing trial integrity, presentation, or publication of the results

DMC indicates data monitoring committee.

All DMC members have to be in agreement before a recommendation to stop early can be made to the executive committee. Some may argue that if the data are not compelling enough to achieve agreement among all DMC members, then the study should not be stopped early. The charter should contain predefined statistical stopping guidelines. The statistical stopping guideline chosen for a trial should assure that overwhelming evidence is required to stop the trial early for benefit. Appropriate statistical methods exist to guide decisions to stop for benefit.\(^{40-42}\) Clinical judgment must also be integrated into the decision-making process, along with formal statistical rules. Several previously described examples (CHARM, SHIFT) illustrate situations in which clinical judgment outweighed the fact that predefined stopping boundaries had been crossed. DMC members who are knowledgeable and experienced with these methodologies and their respective strengths and weaknesses should be involved in the selection of the stopping guideline.

In addition to these guidelines, the DMC charter should also give consideration to the extent of evidence that is needed (eg, the number of events that will be required and/or the minimum length of follow-up required) before early termination for benefit can be recommended, as well as the level of significance that will be required to have sufficient confidence in the results. Fewer interim looks that occur later in the follow-up period is one approach that may minimize the potential of observing a random high.\(^{10,13}\) The DMC should exercise its judgment on the basis of the stopping guidelines as well as the totality of evidence, both internal and external to the trial.

From an operational point of view, we suggest that trial protocols include a section outlining study procedures that would apply if the trial is stopped early. Potential items covered in this section may include:

- Procedures to communicate with subjects and rapidly schedule close-out visits
- Procedures to rapidly communicate with ethics committees
- Plan to switch patients to active drug (if effective)
- Language in the original consent to cover the possibility of an open-label active-drug extension phase, with the goal of minimizing the lag time in transitioning subjects to effective therapy
- Procedures to rapidly collect outstanding data and resolve data queries
- Rapid collection of source documents needed to finalize event adjudication
- Process to rapidly address any pending adjudication assessments
- Process for expedited adjudication while maintaining high quality standards
- Confidentiality plan that allows release of necessary information to treat patients appropriately without jeopardizing trial integrity, presentation, or publication of the results

Conclusions

The decision whether to recommend that a trial stop early for benefit is a major challenge for any DMC. Maintaining the integrity of the trial and obtaining precise final results must be balanced against the risks for patients who are randomly assigned to an apparently inferior treatment and the need to rapidly disseminate evidence supporting a treatment benefit to the broader community. The suggestions documented here may help DMCs anticipate and plan for the challenges they may face when considering whether or not to stop a trial early. This dialogue among clinical researchers, scientists, regulators, and statisticians should continue regarding the evidential, statistical, and practical issues that arise in data monitoring and interim analyses so that the overall patients’ best interests can be served.

Acknowledgments

The following individuals were speakers or panelists discussing this topic at the December 2010, 7th Global Cardiovascular Clinical Trials Forum, Paris, France: Jay Cohn, MD; Lawrence Fine, MD; Gerasimos Filippatos, MD; Mihai Gheorghiade, MD; David Gordon, PhD; Yasser Khder, MD; Wolfgang Koenig, MD; Alice Mascette, MD; Roxana Mehran, MD; Marco Metra, MD; Marc Pfeffer, MD, PhD; Harry Shi, MD; Christian Torp-Pedersen, MD; John Vincent, MD; Hans Wedel, MD; David Whellan, MD; and Andrew Zalewski, MD.

Sources of Funding

This report was generated from discussions during the 7th Global Cardiovascular Clinical Trials (CVCT) Forum held in Paris, France, in December 2010. CVCT was organized by the Clinical
References


Key Words: clinical trial | clinical trials data monitoring committees | data interpretation | statistical stopping rules for benefit
When to Stop a Clinical Trial Early for Benefit: Lessons Learned and Future Approaches
Faiez Zannad, Wendy Gattis Stough, John J.V. McMurray, Willem J. Remme, Bertram Pitt, Jeffrey S. Borer, Nancy L. Geller and Stuart J. Pocock

Circ Heart Fail. 2012;5:294-302
doi: 10.1161/CIRCHEARTFAILURE.111.965707
Circulation: Heart Failure is published by the American Heart Association, 7272 Greenville Avenue, Dallas, TX 75231
Copyright © 2012 American Heart Association, Inc. All rights reserved.
Print ISSN: 1941-3289. Online ISSN: 1941-3297

The online version of this article, along with updated information and services, is located on the World Wide Web at:
http://circheartfailure.ahajournals.org/content/5/2/294

Permissions: Requests for permissions to reproduce figures, tables, or portions of articles originally published in Circulation: Heart Failure can be obtained via RightsLink, a service of the Copyright Clearance Center, not the Editorial Office. Once the online version of the published article for which permission is being requested is located, click Request Permissions in the middle column of the Web page under Services. Further information about this process is available in the Permissions and Rights Question and Answer document.

Reprints: Information about reprints can be found online at:
http://www.lww.com/reprints

Subscriptions: Information about subscribing to Circulation: Heart Failure is online at:
http://circheartfailure.ahajournals.org//subscriptions/