

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 $\leq 30\%$ ($\leq 35\%$ 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 ($P = 0.008$).¹⁴

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.); Sticars Cardiovascular Research Institute, Rhondda, 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.^{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 β -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.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.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 now 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, $P=0.02$.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 al¹ 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.^{2–4} 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.⁹ 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.⁹ 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.”^{9,10} 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.¹⁹

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.²¹ 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; $P<0.00001$; revascularization or hospitalization for unstable angina: RR, 0.53; 95% CI, 0.40–0.70; $P<0.00001$; 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 .^{22,23}

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 1^{9,14,15,19,27–39} 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; $P = 0.092$).³⁵

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

Trial	Stopping Procedures	End Point for Stopping	Length of Follow-Up	Stopped Early, Yes or No	Final Results	Clinical Implications (Acceptance/Uptake)
Heart failure pharmacological trials						
CONSENSUS ²⁷	No formal stopping rule in place	DMC reviewed data every 3 mo for mortality	Actual: 188 d (mean); planned: 12 mo	Yes	n=253 All-cause mortality: 50 (39%) enalapril vs 68 (54%) placebo, $P=0.003$	Cornerstone of HF management, guideline-recommended therapy
GESICA ²⁸	DMC reviewed after one-third and two-thirds of planned enrollment	All-cause mortality	Actual: 13 mo (mean); planned: 2 y	Yes	n=516 Mortality: 87 amiodarone vs 106 control; RR, 28%; 95% CI, 4–45%, $P=0.024$	Stopped on random high, “regression to truth” with CHF STAT ²⁹
CHF STAT ²⁹	Stopping procedures not described in results report	Not specified	Actual: 45 mo (median) Planned: Minimum 1 y	No	n=674 Mortality: 131 (39%) amiodarone vs 143 (42%) placebo, $P=0.6$	Follow-up >3 × longer, more patients, more events than GESICA
US Carvedilol Trials ³⁰	DMC met periodically to review blinded results, no formal rules for stopping	All-cause mortality for the stratified trial program (4 individual trials with nonfatal primary end points)	Actual: 6.5 mo (median) Planned: 12 mo	Yes	n=1094 Mortality: 22 (3.2%) carvedilol vs 31 (7.8%) placebo; $P<0.001$	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
CIBIS-II ¹⁹	Peto rule $P<0.001$ for all-cause mortality	All-cause mortality	Actual: 1.3 y (mean) Planned: 2 y	Yes	n=2647 Mortality: 156 (11.8%) bisoprolol vs 228 (17.3%) placebo; HR, 0.66; 95% CI, 0.54–0.8, $P<0.0001$	Results confirmed observations in US Carvedilol
MERIT-HF ³¹	Asymmetric group-sequential procedure, with cumulative probability of stopping early of 0.0036	All-cause mortality	Actual: 1 y (mean) Planned: 2.4 y	Yes	n=3991 Mortality: 145 (7.2%) metoprolol XL vs 217 (11%) placebo; RR, 0.66; 95% CI, 0.53–0.81, $P=0.0062$	Results consistent with totality of evidence
COPERNICUS ³²	Truncated O'Brien-Fleming type boundary computed with the Lan-DeMets procedure	All-cause mortality	Actual: 10.4 mo (mean) Planned: event driven to 900 deaths	Yes	n=2289 Mortality: 130 (11.4%) carvedilol vs 190 (18.5%) placebo; RR, 35%; 95% CI, 19–48%, $P=0.0014$	Results consistent with totality of evidence
RALES ¹⁵	Group sequential monitoring plan with Lan-DeMets stopping boundary and an O'Brien Fleming spending function	All-cause mortality	Actual: 24 mo (mean) Planned: 3 y	Yes	n=1663 Mortality: 284 (35%) spironolactone vs 386 (46%) placebo; RR, 0.7; 95% CI, 0.6–0.82, $P<0.001$	Wide acceptance, even in patients not reflected among study population
A-HeFT ³³	Lan DeMets sequential boundaries	All-cause mortality (differed from primary composite end point of all-cause death, HF hospitalization, and change in quality of life)	Actual: 10 mo (mean) Planned: 18 mo	Yes	n=1050 Mortality: 32 (6.2%) Bidil vs 54 (10.2%) placebo; HR, 0.57, $P=0.01$	Accepted and incorporated as guideline-recommended therapy, although uptake has been suboptimal
CHARM ^{9,34}	Haybittle-Peto rule, requiring 2-sided $P<0.001$ for the overall program using a log-rank test stratified by trial; for interim analyses within 18 mo of randomization, $P<0.0001$ was required	All-cause mortality (differed from primary end point of overall trial, which was CV death or HF hospitalization)	Actual: 37.7 mo (median) Planned: 2 y (minimum)	No	n=7599 Primary (CV death or HF hospitalization); adjusted HR, 0.82; 95% CI, 0.77–0.91, $P<0.0001$ All-cause mortality; adjusted HR, 0.90; 95% CI, 0.83–1.00, $P=0.032$	Guideline-recommended therapy in ACE inhibitor-intolerant patients or as add-on therapy

(Continued)

Table 1. Continued

Trial	Stopping Procedures	End Point for Stopping	Length of Follow-Up	Stopped Early, Yes or No	Final Results	Clinical Implications (Acceptance/Uptake)
EMPHASIS ¹⁴	Adaptation of the Haybittle-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 $P < 0.001$	CV death or HF hospitalization	Actual: 21 mo (median) Planned: 48 mo	Yes	n=2737 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, $P < 0.001$ All-cause mortality: 171 (12.5%) eplerenone vs 213 (15.5%) placebo, adjusted HR, 0.76; 95% CI, 0.62–0.93, $P = 0.008$	Major guidelines undergoing revision to include eplerenone for patients with mild HF symptoms
SHIFT ³⁵	Peto procedure $P = 0.001$ at each of 2 interim analyses	CV death or HF hospitalization	22.9 mo (median)	No	n=6505 CV death or HF hospitalization: 793 (24%) ivabradine vs 937 (29%) placebo; HR, 0.82; 95% CI, 0.75–0.90, $P < 0.0001$ All-cause mortality: 503 (16%) ivabradine vs 552 (17%) placebo; HR, 0.90; 95% CI, 0.80–1.02, $P = 0.092$	
Heart failure device trials						
MADIT ³⁶	Triangular sequential design with prespecified stopping boundaries	All-cause mortality	Actual: 27 mo Planned: 5 y	Yes	n=196 Mortality: 15/95 defibrillator vs 39/101 conventional treatment; HR, 0.46; 95% CI, 0.26–0.82, $P = 0.009$	Accepted, guideline-recommended device therapy
MADIT II ³⁷	Triangular sequential design with prespecified stopping boundaries	All-cause mortality	Actual: 20 mo (mean) Planned: 2 y	Yes	n=1232 Mortality: 105 (14.2%) ICD vs 97 (19.8%) no ICD; HR, 0.69; 95% CI, 0.51–0.93, $P = 0.016$	Accepted, guideline-recommended device therapy
COMPANION ³⁸	O'Brien Fleming implemented by Lan and DeMets	All-cause death or all-cause hospitalization	Actual: 14.8–16.5 mo (median) Planned: Event-driven (1000 events)	Yes	n=1520 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, $P = 0.003$	Accepted, guideline-recommended device therapy
MADIT-CRT ³⁹	Wang-Tsiatis group sequential design; 20 interim analyses with approximately 35 events each	All-cause death or nonfatal HF events	Actual: 2.4 y (mean) Planned: 2 y (minimum)	Yes	n=1820 Deaths or HF events: 187 (17.2%) CRT-ICD vs 185 (25.3%) ICD only; HR, 0.66; 95% CI, 0.52–0.84, $P = 0.001$	Guideline recommendations have been extended to NYHA class II on the basis of this trial and other recently completed trials

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.

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 rele-

vant 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 2. Considerations for Stopping a Trial Early

Charter considerations (with respect to decisions to stop early)
Role of DMC as advisory or decision-making body and to whom it should report
Need for full agreement of DMC for early stopping recommendations. If full agreement not required, proportion of members that must be in agreement
Predefined statistical stopping boundaries
Clinical parameters that will be considered in decision
Number of events required
Length of follow-up required
Level of significance required
Totality of evidence
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 Trialists Forum, Paris, France: Jay Cohn, MD; Lawrence Fine, MD; Gerasimos Filippatos, MD; Mihai Gheorghide, 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 Zaleski, MD.

Sources of Funding

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

Investigation Center (CIC) Inserm, CHU, and University Henri Poincaré of Nancy, France, and funded by an unrestricted educational grant from Association de Recherche et d'Information en Cardiologie (ARISC), a non-profit educational organization, in Nancy, France. ARISC had no involvement in writing the manuscript or decisions related to submitting for publication.

Disclosures

Dr Zannad was supported by Pfizer, Inc (Steering Committee); Dr Gattis Stough: INSERM, Centre d'Investigation Clinique, Centre Hospitalier Universitaire, Nancy, France (travel expense reimbursement to attend CVCT 2010); professional/project management/administrative time related to preparation of this report); Dr McMurray: Pfizer (grants, travel expenses, consulting fee related to EMPHASIS-HF); Dr Borer: INSERM, Centre d'Investigation Clinique, Centre Hospitalier Universitaire, Nancy, France (travel expense reimbursement to attend CVCT 2010); Servier, Pfizer, BioMarin, Roche, Novartis, Takeda (drug development consulting, membership on DMCs and event adjudication entities); Servier (speaking); Servier (payment for manuscript preparation); BioMarin (stock/stock options); Dr Geller: INSERM, Centre d'Investigation Clinique, Centre Hospitalier Universitaire, Nancy, France (travel expense reimbursement to attend CVCT 2010); National Heart, Lung, and Blood Institute, Bethesda, MD (employer); Francis-Taylor (royalties from 2005 textbook *Advanced Topics in Clinical Biostatistics*); American Statistical Association (professional travel as 2011 President).

References

- Bassler D, Briel M, Montori VM, Lane M, Glasziou P, Zhou Q, Heels-Ansdell D, Walter SD, Guyatt GH, Flynn DN, Elamin MB, Murad MH, Abu Elnour NO, Lampropoulos JF, Sood A, Mullan RJ, Erwin PJ, Bankhead CR, Perera R, Ruiz CC, You JJ, Mulla SM, Kaur J, Nerenberg KA, Schunemann H, Cook DJ, Lutz K, Ribic CM, Vale N, Malaga G, Akl EA, Ferreira-Gonzalez I, Alonso-Coello P, Urrutia G, Kunz R, Bucher HC, Nordmann AJ, Raatz H, da Silva SA, Tuche F, Strahm B, Djulbegovic B, Adhikari NK, Mills EJ, Gwadrý-Sridhar F, Kirpalani H, Soares HP, Karanickolas PJ, Burns KE, Vandvik PO, Coto-Yglesias F, Chripim PP, Ramsay T. Stopping randomized trials early for benefit and estimation of treatment effects: systematic review and meta-regression analysis. *JAMA*. 2010;303:1180–1187.
- Korn EL, Freidlin B, Mooney M. Bias and trials stopped early for benefit. *JAMA*. 2010;304:157–158.
- Ellenberg SS, DeMets DL, Fleming TR. Bias and trials stopped early for benefit. *JAMA*. 2010;304:158–159.
- Goodman S, Berry D, Wittes J. Bias and trials stopped early for benefit. *JAMA*. 2010;304:157–158.
- Korn EL, Freidlin B, Mooney M. Stopping or reporting early for positive results in randomized clinical trials: the National Cancer Institute Cooperative Group experience from 1990 to 2005. *J Clin Oncol*. 2009;27:1712–1721.
- Trotta F, Apolone G, Garattini S, Tafuri G. Stopping a trial early in oncology: for patients or for industry? *Ann Oncol*. 2008;19:1347–1353.
- Mueller PS, Montori VM, Bassler D, Koenig BA, Guyatt GH. Ethical issues in stopping randomized trials early because of apparent benefit. *Ann Intern Med*. 2007;146:878–881.
- Montori VM, Devereaux PJ, Adhikari NK, Burns KE, Eggert CH, Briel M, Lacchetti C, Leung TW, Darling E, Bryant DM, Bucher HC, Schunemann HJ, Meade MO, Cook DJ, Erwin PJ, Sood A, Sood R, Lo B, Thompson CA, Zhou Q, Mills E, Guyatt GH. Randomized trials stopped early for benefit: a systematic review. *JAMA*. 2005;294:2203–2209.
- Pocock S, Wang D, Wilhelmsen L, Hennekens CH. The data monitoring experience in the Candesartan in Heart Failure Assessment of Reduction in Mortality and morbidity (CHARM) program. *Am Heart J*. 2005;149:939–943.
- Pocock SJ. When (not) to stop a clinical trial for benefit. *JAMA*. 2005;294:2228–2230.
- Pocock S, Wilhelmsen L, Dickstein K, Francis G, Wittes J. The data monitoring experience in the MOXCON trial. *Eur Heart J*. 2004;25:1974–1978.
- Psaty BM, Rennie D. Stopping medical research to save money: a broken pact with researchers and patients. *JAMA*. 2003;289:2128–2131.
- Pocock S, White I. Trials stopped early: too good to be true? *Lancet*. 1999;353:943–944.
- Zannad F, McMurray JJ, Krum H, van Veldhuisen DJ, Swedberg K, Shi H, Vincent J, Pocock SJ, Pitt B. Eplerenone in patients with systolic heart failure and mild symptoms. *N Engl J Med*. 2011;364:11–21.
- Pitt B, Zannad F, Remme WJ. The effect of spironolactone on morbidity and mortality in patients with severe heart failure: Randomized Aldactone Evaluation Study Investigators. *N Engl J Med*. 1999;341:709–717.
- Pitt B, Williams G, Remme W, Martinez F, Lopez-Sendon J, Zannad F, Neaton J, Roniker B, Hurley S, Burns D, Bittman R, Kleiman J. The EPHEsus trial: eplerenone in patients with heart failure due to systolic dysfunction complicating acute myocardial infarction: Eplerenone Post-AMI Heart Failure Efficacy and Survival Study. *Cardiovasc Drugs Ther*. 2001;15:79–87.
- Yusuf S, Sleight P, Pogue J, Bosch J, Davies R, Dagenais G. Effects of an angiotensin-converting-enzyme inhibitor, ramipril, on cardiovascular events in high-risk patients: the Heart Outcomes Prevention Evaluation Study Investigators. *N Engl J Med*. 2000;342:145–153.
- Dahlof B, Sever PS, Poulter NR, Wedel H, Beevers DG, Caulfield M, Collins R, Kjeldsen SE, Kristinsson A, McInnes GT, Mehlsen J, Nieminen M, O'Brien E, Ostergren J. Prevention of cardiovascular events with an antihypertensive regimen of amlodipine adding perindopril as required versus atenolol adding bendroflumethiazide as required, in the Anglo-Scandinavian Cardiac Outcomes Trial-Blood Pressure Lowering Arm (ASCOT-BPLA): a multicentre randomised controlled trial. *Lancet*. 2005;366:895–906.
- CIBIS II Investigators. The Cardiac Insufficiency Bisoprolol Study II (CIBIS-II): a randomised trial. *Lancet*. 1999;353:9–13.
- Ridker PM, Danielson E, Fonseca FA, Genest J, Gotto AM Jr, Kastelein JJ, Koenig W, Libby P, Lorenzatti AJ, MacFadyen JG, Nordestgaard BG, Shepherd J, Willerson JT, Glynn RJ. Rosuvastatin to prevent vascular events in men and women with elevated C-reactive protein. *N Engl J Med*. 2008;359:2195–2207.
- Kaul S, Morrissey RP, Diamond GA. By Jove! What is a clinician to make of JUPITER? *Arch Intern Med*. 2010;170:1073–1077.
- Ridker PM, Glynn RJ. The JUPITER Trial: responding to the critics. *Am J Cardiol*. 2010;106:1351–1356.
- Ridker PM. The JUPITER trial: results, controversies, and implications for prevention. *Circ Cardiovasc Qual Outcomes*. 2009;2:279–285.
- Cohn JN, Goldstein SO, Greenberg BH, Lorell BH, Bourge RC, Jaski BE, Gottlieb SO, McGrew F III, DeMets DL, White BG. A dose-dependent increase in mortality with vesnarinone among patients with severe heart failure: Vesnarinone Trial Investigators. *N Engl J Med*. 1998;339:1810–1816.
- Pitt B, Segal R, Martinez FA, Meurers G, Cowley AJ, Thomas I, Deedwania PC, Ney DE, Snively DB, Chang PI. Randomised trial of losartan versus captopril in patients over 65 with heart failure (Evaluation of Losartan in the Elderly Study, ELITE). *Lancet*. 1997;349:747–752.
- Pitt B, Poole-Wilson PA, Segal R, Martinez FA, Dickstein K, Camm AJ, Konstam MA, Riegger G, Klinger GH, Neaton J, Sharma D, Thyagarajan B. Effect of losartan compared with captopril on mortality in patients with symptomatic heart failure: randomised trial: the Losartan Heart Failure Survival Study ELITE II. *Lancet*. 2000;355:1582–1587.
- CONSENSUS Trial Study Group. Effects of enalapril on mortality in severe congestive heart failure: results of the Cooperative North Scandinavian Enalapril Survival Study (CONSENSUS). *N Engl J Med*. 1987;316:1429–1435.
- Doval HC, Nul DR, Grancelli HO, Perrone SV, Bortman GR, Curiel R. Randomised trial of low-dose amiodarone in severe congestive heart failure: Grupo de Estudio de la Sobrevida en la Insuficiencia Cardiaca en Argentina (GESICA). *Lancet*. 1994;344:493–498.
- Singh SN, Fletcher RD, Fisher SG, Singh BN, Lewis HD, Deedwania PC, Massie BM, Colling C, Lazzari D. Amiodarone in patients with congestive heart failure and asymptomatic ventricular arrhythmia: Survival Trial of Antiarrhythmic Therapy in Congestive Heart Failure. *N Engl J Med*. 1995;333:77–82.
- Packer M, Bristow MR, Cohn JN, Colucci WS, Fowler MB, Gilbert EM, Shusterman NH. The effect of carvedilol on morbidity and mortality in patients with chronic heart failure: US Carvedilol Heart Failure Study Group. *N Engl J Med*. 1996;334:1349–1355.
- MERIT-HF Investigators. Effect of metoprolol CR/XL in chronic heart failure: Metoprolol CR/XL Randomised Intervention Trial in Congestive Heart Failure (MERIT-HF). *Lancet*. 1999;353:2001–2007.
- Packer M, Coats AJ, Fowler MB, Katus HA, Krum H, Mohacsi P, Rouleau JL, Tendera M, Castaigne A, Roecker EB, Schultz MK, DeMets DL. Effect of carvedilol on survival in severe chronic heart failure. *N Engl J Med*. 2001;344:1651–1658.

33. Taylor AL, Ziesche S, Yancy C, Carson P, D'Agostino R Jr, Ferdinand K, Taylor M, Adams K, Sabolinski M, Worcel M, Cohn JN. Combination of isosorbide dinitrate and hydralazine in blacks with heart failure. *N Engl J Med*. 2004;351:2049–2057.
34. Pfeffer MA, Swedberg K, Granger CB, Held P, McMurray JJ, Michelson EL, Olofsson B, Ostergren J, Yusuf S, Pocock S. CHARM Investigators and Committees: Effects of candesartan on mortality and morbidity in patients with chronic heart failure: the CHARM-Overall programme. *Lancet*. 2003;362:759–766.
35. Swedberg K, Komajda M, Bohm M, Borer JS, Ford I, Dubost-Brama A, Lerebours G, Tavazzi L. Ivabradine and outcomes in chronic heart failure (SHIFT): a randomised placebo-controlled study. *Lancet*. 2010;376:875–885.
36. Moss AJ, Hall WJ, Cannom DS, Daubert JP, Higgins SL, Klein H, Levine JH, Saksena S, Waldo AL, Wilber D, Brown MW, Heo M. Improved survival with an implanted defibrillator in patients with coronary disease at high risk for ventricular arrhythmia: Multicenter Automatic Defibrillator Implantation Trial Investigators. *N Engl J Med*. 1996;335:1933–1940.
37. Moss AJ, Zareba W, Hall WJ, Klein H, Wilber DJ, Cannom DS, Daubert JP, Higgins SL, Brown MW, Andrews ML. Prophylactic implantation of a defibrillator in patients with myocardial infarction and reduced ejection fraction. *N Engl J Med*. 2002;346:877–883.
38. Bristow MR, Saxon LA, Boehmer J, Krueger S, Kass DA, De MT, Carson P, DiCarlo L, DeMets D, White BG, DeVries DW, Feldman AM. Cardiac-resynchronization therapy with or without an implantable defibrillator in advanced chronic heart failure. *N Engl J Med*. 2004;350:2140–2150.
39. Moss AJ, Hall WJ, Cannom DS, Klein H, Brown MW, Daubert JP, Estes NA III, Foster E, Greenberg H, Higgins SL, Pfeffer MA, Solomon SD, Wilber D, Zareba W. Cardiac-resynchronization therapy for the prevention of heart-failure events. *N Engl J Med*. 2009;361:1329–1338.
40. Pocock SJ. Current controversies in data monitoring for clinical trials. *Clin Trials*. 2006;3:513–521.
41. Schulz KF, Grimes DA. Multiplicity in randomised trials II: subgroup and interim analyses. *Lancet*. 2005;365:1657–1661.
42. O'Brien P. Data and safety monitoring. In: Armitage P, Colton T, eds. *Encyclopedia of Biostatistics*. Chichester, England: John Wiley & Sons; 1998:1058–1066.
43. Borer JS, Gordon DJ, Geller NL. When should data and safety monitoring committees share interim results in cardiovascular trials? *JAMA*. 2008;299:1710–1712.
44. Fox K, Ford I, Steg PG, Tendera M, Ferrari R. Ivabradine for patients with stable coronary artery disease and left-ventricular systolic dysfunction (BEAUTIFUL): a randomised, double-blind, placebo-controlled trial. *Lancet*. 2008;372:807–816.
45. DAMOCLES Study Group. A proposed charter for clinical trial data monitoring committees: helping them to do their job well. *Lancet*. 2005;365:711–722.

KEY WORDS: clinical trial ■ clinical trials data monitoring committees ■ data interpretation ■ statistical stopping rules for benefit