Development of Therapeutics for Heart Failure: Expedited Commentary

Unbelievable Folly of Clinical Trials in Heart Failure
The Inconvenient Truth About How Investigators and Guidelines Weigh Evidence

Milton Packer, MD

Despite significant advances in the treatment of chronic heart failure during the past 30 years, there is a broad consensus that new treatments are needed because morbidity and mortality in this disorder remain unacceptably high.1 Nevertheless, the development of new drugs for chronic heart failure has waned during the past 20 years; the quantity of novel agents in clinical testing and the number of well-designed clinical trials have diminished to a small fraction of that seen in the 1980s and 1990s.2 Some have suggested that investment in heart failure has declined because the primary mechanisms of the disorder are poorly understood, because the promising results of phase II trials are frequently not confirmed in large-scale definitive studies, and because the large-scale clinical trials needed for regulatory approval have often become prohibitively expensive.2,3 We know that clinical research in heart failure is extremely challenging, and so we beg innovators in the pharmaceutical and device industries to be patient, and we plead with sponsors to invest in large-scale investigations because the need is so great. Yet, when we critically examine how we react to data, it may not be too surprising that meaningful clinical trials in heart failure are becoming increasingly scarce.2 Do cardiologists act as if the results of clinical trials in heart failure really matter?

Begging Forgiveness for Our Failures
Most heart failure trials yield results that are disappointing or difficult to interpret. Yet, such studies would nevertheless be enormously useful if we learned from these setbacks and moved on. But do we? At the 2015 Scientific Sessions of the American Heart Association, every randomized trial in heart failure presented in the late-breaking sessions yielded distressing results. A phase II trial of the soluble guanylate cyclase stimulator vericiguat failed to meet its primary end point, even though the change in the biomarker that the investigators selected was an easy surrogate target.4 Nevertheless, based on exploratory analyses suggesting some effect at high doses, the investigators called for further trials. A trial reported that therapeutic doses of oral nitrates in patients with heart failure and a preserved ejection fraction had an adverse effect on daily activity with no benefit on quality of life or levels of N-terminal proBNP.5 Despite this disappointment, the investigators suggested that alternative nitrate regimens might prove to be useful and proposed that they warranted further evaluation. A trial that evaluated the effects of the GLP-1 agonist lixisenatide in patients with heart failure and a reduced ejection fraction yielded no benefits on the clinical symptoms and course of the patients and suggested the possibility of adverse effects on renal function and an increased risk of death or hospitalization for heart failure.6 Nevertheless, the investigators held out the possibility that the drug might still be useful in patients with less advanced disease. A trial reported that the β3 adrenergic receptor agonist mirabegron failed to produce any of its expected benefits, but the presenters concluded that more trials were needed based on a finding of borderline significance in a post hoc analysis.7 Undermining Our Rare Successes
Given the low chances for success, we might think the cardiology community would celebrate the rare occasions when we are able to report a favorable and meaningful result in a large-scale clinical trial. However, it is not clear that we are genuinely capable of rejoicing even under the happiest circumstances. When the results of a positive trial are presented at a scientific meeting, clinical trial experts relish in identifying perceived weaknesses in the study design or analysis, even if such deficiencies are slight and fail to meaningfully alter the interpretation of results. Undue emphasis is often placed on subgroups that can differ by the play of chance alone.8 Trials carried out in certain geographical regions are often criticized for their lack of inclusion of special populations, even though there may be little reason to think that special groups might exhibit responses that differ from those who were enrolled in the study. Experts criticize the trials for not measuring the effect on some surrogate end point, failing to understand that physiological measures and biomarkers are intended to predict the results of outcome trials rather than replace them.9 Trials with particularly favorable results may be terminated early for appropriate ethical considerations, only to have the early closure used as a reason for suggesting that the magnitude of the observed effects is exaggerated or to have the shortening of follow-up be used to raise the false issue that the trial failed to yield long-term safety data.10 Could a new

Received January 22, 2016; accepted February 5, 2016.
From the Baylor Heart and Vascular Institute, Baylor University Medical Center, Dallas, TX.
Correspondence to Milton Packer, MD, Baylor Heart and Vascular Institute, Baylor University Medical Center, 621 N. Hall St, Dallas, TX 75226. E-mail milton.packer@baylorhealth.edu
(Circ Heart Fail. 2016;9:e002837. DOI: 10.1161/CIRCHEARTFAILURE.116.002837.)
© 2016 American Heart Association, Inc.
Circ Heart Fail is available at http://circheartfailure.ahajournals.org DOI: 10.1161/CIRCHEARTFAILURE.116.002837

drug that prolongs life in heart failure conceivably cause some rare dermatologic disease after 15 years of use? The mere possibility of any unwarranted uncertainty is presented as a valid reason for clinicians to exercise caution.

I am not suggesting that the heart failure community embrace every trial that reports some nominally significant $P$ value for every promising post hoc analysis. I am certainly not proposing that trials that seem to meet their primary end point should be viewed as providing undeniable truths that should be immediately trumpeted to physicians throughout the world. Yet once every several years, our colleagues do manage to design, conduct, and present results that are robust and meaningful. Such trials are really worthwhile, and they will change life for patients with heart failure for the good. Yet, the feeding frenzy that they generate is remarkable to watch. There should be no doubt: if you are skeptical about every clinical trial that you hear about, you will not find yourself on the record having supported a study that subsequently becomes problematic. Yet, is it intellectually useful to invariably provide a cynical view? As a member of and consultant to the FDA Advisory Committee for Cardiac and Renal Drug Products for nearly 30 years, I am reminded of certain members who voted against the approval of every single drug they evaluated, even if the evidence for a favorable benefit-to-risk relation was overwhelming. By assuming the same predictable position, they made clear that their opinions were uninformed by the data. I can appreciate that taking a negative position on every set of data is the safest political position that anyone can assume, but such a position is not particularly useful for patients.

**Believing in the Absurdity of the 1 Versus 2 Trial Rule**

Assuming that a trial is actually well-designed and well-executed and has yielded robust results, do we accept its findings and make recommendations to our colleagues? Given the great needs of our patients and our own pleas for new therapies, one might expect a truly effective treatment to be rapidly approved by regulatory agencies and quickly incorporated into guidelines issued from organizations charged with providing leadership in this area. However, building the consensus required to issue clinical recommendations is a political—not a scientific—process, and like all political processes, it is messy.

Guidelines committees are not well positioned to tackle design biases in a clinical trial, audit the data integrity of available studies, examine their statistical validity of conclusions or perform a thorough benefit-risk assessment of any treatment. These tasks are carried out with a high degree of professionalism and integrity by regulatory agencies, which have access to raw data and are able to perform hundreds of analyses that are capable of challenging the robustness of a trial’s findings. Understandably, those with the greatest expertise in the design and analysis of clinical trials are generally excluded from the guidelines process because of conflicts of interest. So we should ask: once the regulators have completed their task and have decided to approve a drug for general use, what do leaders in cardiology do? Often, they sit in silence. What reason do they offer for their caution? They refer to the need for replication, the need for 2 trials, the need to bring the level of uncertainty down to an imperceptible level; yet, even when such circumstances are eventually fulfilled, they often fail to pay attention to the evidence.

I am amazed by the current fondness that the guidelines process has for creating a hierarchy of recommendations that purports to summarize the strength of conviction that supports a clinical recommendation. The American College of Cardiology/American Heart Association (ACC/AHA) guidelines process distinguishes between a class I and class II recommendation based on the size of the treatment effect, with a class I recommendation indicating that the magnitude of the effect is meaningful and that the benefit clearly outweighs the risk; this is generally the case in large-scale trials showing a significant mortality reduction. At the same time, the guidelines process also attempts to provide either a class A or class B recommendation based on the level of certainty that the effect (regardless of size) is real. According to the conventional approach, class A evidence is based on multiple trials or meta-analyses, whereas class B evidence is based on a single-randomized trial or nonrandomized studies. Such an approach sounds appealing, but it ignores the fact that the primary basis of determining the certainty of a finding is the quality of the data, rather than some arbitrary level of the quantity of data. If only quantity counted, 2 small trials with problematic designs and borderline statistically significant results without internal consistency would be considered to yield evidence that would be deemed more reliable than one compelling large-scale trial, which showed a high level of certainty (with a very small $P$ value) and substantial internal consistency across subgroups and across end points. The creation of a hierarchy based on quantity rather than quality makes no sense—clinically, statistically, or ethically.

Take, for example, the current guideline recommendation for the use of angiotensin receptor blockers for the treatment of chronic heart failure, which currently carries a class IA recommendation (meaning the effect is large and clinically meaningful, and we are certain about its benefit). According to the 2013 ACC/AHA guidelines, angiotensin receptor blockers are recommended to reduce morbidity and mortality in patients with heart failure with current or previous symptoms who are intolerant of an angiotensin-converting enzyme (ACE) inhibitor. Such a vigorous recommendation seems odd, however, because there is only 1 trial that showed a benefit of an angiotensin receptor blocker in ACE intolerant patients with chronic heart failure. That trial did not enroll asymptomatic patients; and the trial did not demonstrate a significant effect on mortality, which was reduced by <10%. One additional trial (Valsartan Heart Failure Trial [Val-HeFT]), which demonstrated no overall mortality benefit with the angiotensin receptor blocker valsartan in chronic heart failure, reported some benefit in a small retrospectively defined subgroup of patients not taking ACE inhibitors (but not demonstrably intolerant of them), but this small subgroup showed no significant reduction in cardiovascular mortality in an analysis that was based on <80 events. To add to the confusion, the guidelines document emphasizes that ACE inhibitors may have an advantage over angiotensin receptor blockers, which is related to the accumulation of kinins rather than to the suppression of angiotensin II formation, and yet, it cites the results of a
trial—not carried out in chronic heart failure—which reported no difference between the effects of an ACE inhibitor and an angiotensin receptor blocker. The document considers angiotensin receptor blockers (which have minimal effects on kinins) to be a reasonable alternative to ACE inhibitors; yet, it fails to make any reference to the only 2 existing direct comparator trials, both of which reported that ACE inhibitors may indeed be superior to an angiotensin receptor blockers in their effects on cardiovascular mortality.16,17 Does this makes sense, given the fact that the guidance for angiotensin receptor blockers represents a class IA recommendation?

The current ACC/AHA guideline for the use of the combination of hydralazine and isosorbide dinitrate is similarly puzzling. The class IA recommendation—that the drug combination should be used to reduce morbidity and mortality in black patients—is based on the results of a single trial,18 whose primary end point was not morbidity and mortality, but instead, a previously unused composite, whose clinical and statistical validity was criticized when the findings were discussed at a public advisory committee. In the original protocol, a positive finding of a mortality reduction in the trial was considered so unlikely that the Data and Safety Monitoring Committee set up no prespecified rules for early stopping for a survival benefit. A second trial, presumably used to justify a class IA recommendation, was a post hoc analysis of a tiny subgroup of the Vasodilator Heart Failure (V-HeFT) I trial,19 which enrolled no patients receiving any of the neurohor- monal antagonists currently mandated for clinical use and for which the conventional criterion for a potentially meaningful statistical interaction was not met. Ironically, the combination of hydralazine and isosorbide dinitrate actually proved to be nearly inferior to an ACE inhibitor on survival in the patients with mild-to-moderate heart failure enrolled in the V-HeFT II trial,20 even though the effects of an ACE inhibitor on mortality in such patients are known to be modest (when compared with placebo).21 Interestingly, the effect of hydralazine and isosorbide dinitrate on the risk of death and heart failure hospitalizations was not a prespecified end point in either V-HeFT I or V-HeFT II,20,22 and the combined data from V-HeFT I and V-HeFT II were sufficiently confusing that the FDA declined to approve the combination for the treatment of chronic heart failure when the initial application for its commercial use was made in 1997.23

At the same time, the guideline process has been silent on 2 drugs that were approved for the treatment of heart failure in the United States >6 months ago, ivabradine and sacubitril/valsartan. I was one of the 2 principal investigators of a trial,24 which demonstrated the survival benefits of adding a neprilysin inhibitor to an inhibitor of the renin–angiotensin system. Despite favorable opinions from unconflicted external reviewers supporting its adoption as first-line therapy,25–27 the results of the study have been described as providing evidence from only a single-randomized trial. For some,28,29 for sacubitril/valsartan to receive a class IA recommendation, the findings of the mortality trial with the drug would need to be replicated, although it is difficult to understand how this could be ethically accomplished, especially it were only to serve some artificial administrative purpose. Of course, one could recognize that the strength of evidence in the sacubitril/valsartan trial (P=0.000004 for the primary end point and P=0.00008 for cardiovascular mortality) provides assurance of robustness equal to that of 4 independently conducted trials, each showing a reduction in cardiovascular mortality with a P value of <0.05. Alternatively, one could look at favorable analyses in 2 trials with an earlier angiotensin neprilysin inhibitor ( omapatrilat), both of which could be cited to support a beneficial effect of this approach on the combined risk of cardiovascular death and hospitalization.30,31 However, I would ask: do we really need to go through such contrivances simply to allow our colleagues a measure of emotional comfort?

Understanding That Our Folly Has Consequences

Some might argue that I am placing too much importance on the guidelines process. Can we not trust that the truth will prevail in the end? Do we not understand that (by their nature) heart failure cardiologists are conservative, and we accept change with great reluctance? Can we not be confident that—over long periods of time—drugs will find their proper place in therapeutics? Can we not think that convincing results in well-designed clinical trials will eventually be heeded? I would like to be optimistic, but I am not so sure. We have had 3 large-scale trials evaluating the combination of hydralazine and isosorbide dinitrate, and it currently carries a class IA recommendation. The drug combination was approved by the FDA in 2005, and at that time, the ACC/AHA guidelines assigned its use a IIA recommendation,32 allowing insurers to decline payment for the drug because of a lack of priority. Interestingly, the ACC/AHA guidelines then elevated the use of the hydralazine–isosorbide dinitrate combination to a class IB recommendation 4 years later (in 200933) and then to a class IA recommendation 4 years after that (in 201334)—although not a single new piece of trial data had emerged about the efficacy and safety of hydralazine and isosorbide dinitrate in the previous 8 years. How did this 8-year delay serve patients? At present, few patients in the United States receive a combination of hydralazine and isosorbide dinitrate at doses shown to be effective in large-scale trials; the slow and ultradeliberate elevation of the guidance on behalf of the use of isosorbide dinitrate and hydralazine has served no useful purpose. However, we have had several multicenter trials that have demonstrated the benefits of digoxin,34,35 which carried a class IA recommendation in the 2001 ACC/AHA guidelines,36 which was downgraded to a class IIA recommendation in the 2009 and 2013 documents11,33—all again in the absence of any new clinical trial evidence. As a result of these downgrades, some physicians have been asked by insurers to complete forms to justify the use of digoxin in patients with chronic heart failure, and consequently, despite its demonstrable effects on clinical status and hospitalizations for heart failure, the use of this extraordinarily inexpensive and well-tolerated drug is dwindling rapidly. The medicine that we practice is based on eminence rather than evidence.

There is a simple way of summarizing the response of the heart failure community to those who have the resources to support our research efforts. Whether our trials succeed or fail, we want sponsors to support more trials. We will always find a
sliver of hope in a totally neutral trial that has failed to meet any of its expectations, and we will always find something lacking in a trial with overwhelmingly robust results. There is no trial that is too negative, and there is no trial that is good enough. Regrettably, if you ask a clinical trialist for the solution to a problem, the answer will always be—a new clinical trial.

At some point in time, whether it is the public or the private sector, someone is bound to see through our pattern of convenience. Regrettably, if you ask a clinical trialist for the solution to a problem, the answer will always be—a new clinical trial. What we really want?

If we do not fulfill our responsibilities as physicians and as academic leaders, sponsors will stop doing clinical trials, because they will have figured out that our purpose for doing them is not to find answers but to have more work to do. Is that what we really want?

Disclosures
Dr Packer has consulted for Admittance, Amgen, AstraZeneca, Bayer, BioControl, Boehringer Ingelheim, Cardiocor, Cardiorentis, Cytokinetics, Daiichi Sankyo, GlaxoSmithKline, Novartis, Takeda, and ZS Pharma.

References


**Key Words:** clinical trials ■ heart failure
Unbelievable Folly of Clinical Trials in Heart Failure: The Inconvenient Truth About How Investigators and Guidelines Weigh Evidence

Milton Packer

Circ Heart Fail. 2016;9:
doi: 10.1161/CIRCHEARTFAILURE.116.002837

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/9/4/e002837