Double Vision

Replicating a Trial Showing a Survival Benefit

Milton Packer, MD

Many physicians boast that there is no reason to read or keep up with the medical literature. Why? If it is true that 1 trial is never enough to change clinical practice, clinicians can always claim that they are waiting for confirmatory evidence and for critical review by regulatory authorities and peer groups. Such thinking greatly simplifies the challenge of maintaining medical knowledge. Practitioners need not become familiar with any scientific study; they need only to wait for guideline statements, which will determine when a critical mass of evidence has been compiled. Sadly, some clinicians will also wait to see what practices and procedures are reimbursed, because this—not scientific evidence—has become a major factor influencing medical care.

This current approach to medical practice is egregiously wrong. The belief that there is no need to be up-to-date on medical evidence—because no observation is valid unless it has been replicated—represents the worst possible excuse for intellectual lethargy and clinical inertia. Furthermore, it allows the practitioner to relegate the process of distilling and deciphering the totality of evidence to others, who may have their own limitations or conflicts. The present state of affairs also fosters efforts to generate replicative evidence by artificial (and often ethically dubious) means, in order to be able to state categorically that an observation has been confirmed.

A HISTORY OF HEART FAILURE TRIALS SHOWING A SURVIVAL BENEFIT

It is scientifically unnecessary and ethically impossible to replicate a trial that shows (with a high degree of confidence) that an intervention prolongs life. Such a situation, however, must be distinguished from the observations of early development trials that report

imbalances in the reported risk of death between 2 treatment groups. When conclusions about drug effects are based on a small number of total events (e.g., <200), treatment differences (although large in magnitude) are determined more by the play of chance that by a true effect of the intervention. The reliability of a finding in an underpowered trial depends not on its nominal p value, but on the narrowness of the confidence intervals, and thus, its likelihood of replication. When a survival benefit is reported in a trial not designed to examine the effect of the intervention on the risk of death, the findings (with respect to survival) are unreliable (because they are based on sparse information) and often, are not confirmed by subsequent trials (Table 1). This is true even though the patient population being studied in the definitive trial is precisely the same as the one that originally was the basis of the reported survival benefit.

This situation, however, differs dramatically from the finding of a survival benefit in a trial that has been specifically designed to evaluate the influence of a new treatment on the risk of death (Table 2). Such trials are rare; during the past 40 years, fewer than 10 trials that have been designed to evaluate the effect of a drug on mortality in patients with chronic heart failure have reported a survival benefit. The number of events supporting the conclusion of a mortality risk reduction in such trials is generally larger than 200 and is substantially greater than the number studied in the early development trials (Table 1).

However, as is seen in Table 2, no effort has ever been made to replicate the results of a definitive finding of a survival benefit in the same population of patients with heart failure that was originally studied. In the case of angiotensin-converting enzyme (ACE) inhibitors, the 3 trials evaluated entirely distinct nonoverlapping groups of patients. Several large-scale trials reported similar effects of different beta-blockers on mortality in chronic heart failure, but the trials were carried out at the same time or were carried out in patients that were excluded from earlier trials. The 3 large-scale outcome trials with mineralocorticoid receptor antagonists studied

TABLE 1 Failure of Replication of a Reported (but Unreliable) Survival Benefit in Randomized Controlled Trials in Chronic Heart Failure

Study Drugs	Trial/First Author	Type of Patients Studied	Total Number of Deaths	Reported % Change in Risk of Death (95% CI)
Vesnarinone vs. placebo	Vesnarinone Study Group	Primarily class III heart failure	46	↓ 62 (28 to 80)
	VEST	Primarily class III heart failure	534	↑ 22 (5 to 42)
Captopril vs. losartan	ELITE-I	Primarily class II-III heart failure	49	↓ 46 (31 to 95)
	ELITE-II	Primarily class II-III heart failure	530	↑ 13 (−5 to 42)
Amlodipine vs. placebo	PRAISE-1	Nonischemic cardiomyopathy, class III-IV	119	↓ 46 (21 to 63)
	PRAISE-2	Nonischemic, cardiomyopathy, class III-IV	1,074	↑ 9 (−8 to 29)
Oxidized autologous blood vs. placebo	Torre-Amione et al.	Primarily class III heart failure	8	Not reported; only p = 0.022 is shown
	ACCLAIM	Primarily class III heart failure	245	↑ 8 (−16 to 39)

See online supplement for references.

ACCLAIM = Advanced Chronic Heart Failure CLinical Assessment of Immune Modulation; CI = confidence interval; ELITE-I = Evaluation of Losartan in the Elderly Study; ELITE-II = Losartan Heart Failure Survival Study II; PRAISE = Prospective Randomized Amlodipine Survival Evaluation; VEST = Vesnarinone Trial.

nonoverlapping patient populations. With respect to the combination of hydralazine and isosorbide dinitrate, the second trial focused on patients who were not studied in the earlier trial (i.e., women and African Americans receiving neurohormonal antagonists).

Hence, in the field of heart failure, there has never been replication of the result of a survival benefit if it was reported in a definitively designed trial. When the benefits of a drug have been convincingly demonstrated in a specific patient population, subsequent trials (if carried out) always focus on studying a distinct and nonoverlapping group of patients. A similar population is targeted for confirmation *only* when the initial report of a survival difference is generated as a result of an unreliable estimate in a trial reporting a small number of events (Table 1). This point is often missed in the development of heart failure guidelines: a Class IA recommendation for a survival benefit is never based on replication of the results with the same drug in the same clinical setting. Instead, a Class IA recommendation for mortality reduction (when based

TABLE 2 Similarity of Survival Benefits in Randomized Controlled Trials When Complementary Groups of Patients With Heart Failure Were Studied

Study Drugs	Trial	Type of Patients Studied	Total Number of Deaths	Reported % Change in Risk of Death (95% CI)
ACE inhibitor vs. placebo	CONSENSUS	Enalapril in class IV heart failure	122	↓ 27 (CI not reported)
	SOLVD-T	Enalapril in class II-III heart failure	962	↓ 16 (5 to 26)
	AIRE	Ramipril in post-MI heart failure	392	↓ 27 (11 to 40)
Beta-blocker vs. placebo	MERIT-HF	Metoprolol, primarily in class II-III heart failure	362	↓ 34 (19 to 47)
	CIBIS II	Bisoprolol, primarily in class II-III heart failure	384	↓ 34 (19 to 46)
	COPERNICUS	Carvedilol, in class IIIB-IV heart failure, EF <25%	320	↓ 35 (19 to 48)
Mineralocorticoid receptor antagonist vs. placebo	RALES	Spironolactone, class III-IV heart failure with recent class IV symptoms	1,054	↓ 30 (18 to 40)
	EMPHASIS-HF	Eplerenone, class II heart failure	384	↓ 22 (5 to 36)
	EPHESUS	Eplerenone in post-MI heart failure	1,032	↓ 15 (4 to 25)
Hydralazine-isosorbide dinitrate combination vs. placebo	V-HeFT	Primarily white men with class II-III heart failure not receiving neurohormonal blockers	192	Not reported; only p = 0.093 is shown
	A-HeFT	African American men and women with class II-III heart failure receiving neurohormonal blockers	86	↓ 43 (CI not reported)

See online supplement for references.

A-HeFT = African-American Heart Failure Trial; AIRE = Acute Infarction Ramipril Efficacy; CI = confidence interval; CIBIS II = The Cardiac Insufficiency Bisoprolol Study II; CONSENSUS = Cooperative North Scandinavian Enalapril Survival Study; COPERNICUS = Carvedilol Prospective Randomized Cumulative Survival Study; EF = ejection fraction; EMPHASIS- HF = Eplerenone in Mild Patients Hospitalization and Survival Study in Heart Failure; EPHESUS = Eplerenone Post-Acute Myocardial Infarction Heart Failure Efficacy and Survival Study; MERIT-HF = Metoprolol CR/XL Randomised Intervention Trial in Congestive Heart Failure; MI = myocardial infarction; RALES = Randomized Aldactone Evaluation Study; SOLVD-T = Studies of Left Ventricular Dysfunction Trial; V-HeFT = Vasoldilator Heart Failure Trial.