Circulation: Heart Failure

EDITORIAL

Embracing the Long Road to Precision Medicine

See Article by Ferreira et al

Julio A. Chirinos, MD, PhD
David E. Lanfear, MD

In classic cardiovascular phase III trials, individuals satisfying a well-defined set of inclusion and exclusion criteria are randomly assigned to receive one of 2 or more treatments in a nonadaptive parallel-arm design. If the trial is positive, for example in favor of treatment A (a new treatment) over treatment B (standard of care), it is concluded that the new treatment benefits such a patient population. A more concrete hypothetical trial may demonstrate that treatment A significantly reduces the composite end point of death or heart failure–related hospitalizations by 20% relative to treatment B. A subjective interpretation that may follow such trial results is that a patient that meets trial criteria will experience a benefit from the treatment (ie, a reduction in adverse outcomes). This interpretation is flawed. In reality, some individuals benefit, some individuals are harmed, and some individuals experience neither benefit nor harm, but the number of individuals who benefit exceeds the number of those who are harmed, such that the net rate of the end point is ≈20% lower in a group of patients receiving treatment A relative to treatment B (for the purposes of our discussion, we will ignore the confidence interval). Explicitly acknowledging this difference in interpretations has increasingly important implications in the era of precision medicine.

Implantable cardioverter defibrillator (ICD) trials can further illustrate these concepts. Consider the SCD-HeFT (Sudden Cardiac Death in Heart Failure Trial),[1] which demonstrated that among patients with heart failure with reduced ejection fraction (HFrEF), New York Heart Association class II or III symptoms, and a left ventricular ejection fraction ≤35%, ICD therapy decreased the risk of death by 23%. Patients who meet indications for ICDs based on well-designed phase III trials may experience infection, bleeding, or mechanical complications as a direct consequence of the implantation, resulting in significant morbidity and even death.[2] If a given patient experiences a fatal complication along with the demonstrable absence of any therapeutic shocks or pacing ever delivered, the presence of harm and absence of benefit in that individual become abundantly clear. Some individuals do not suffer harm but never derive a benefit from the ICD because they never develop fatal ventricular arrhythmias, succumbing instead to competing causes of death (such as progressive heart failure). Yet, some individuals experience (and are successfully rescued from) life-threatening ventricular arrhythmias, and the number of patients in this subgroup exceeds the number of patients who experience the rare fatal complications of ICD implantation. Individuals who are rescued from fatal arrhythmias drive the overall mortality benefit seen in ICD trials but compose a rather small proportion of all subjects enrolled (consider the absolute mortality difference of 29% versus 22% in the placebo and ICD arms of SCD-HeFT).[1] The number needed to treat (in this case, ≈14) can be readily estimated from trial results and is

Key Words: Editorials ■ aldosterone ■ heart failure ■ patients ■ precision medicine

http://circheartfailure.ahajournals.org

a useful concept, but this metric obscures the fact that most patients in the ICD arm of the trial never actually got treated by the device (79% of implanted ICDs not once delivered a shock for rapid ventricular tachycardia or fibrillation in SCD-HeFT).[1] Interestingly, when rapid ventricular tachycardia or fibrillation does occur, ICDs are highly effective. The number needed to treat, thus, did not reflect an inability of ICDs to effectively abort arrhythmic SCD but rather the high proportion of ICDs needed to implant which never actually treated patients for life-threatening ventricular arrhythmias. This resulted from the limited precision in identifying patients at high risk of dying from these arrhythmias (which in this case are the therapeutic target).

Discerning the direct benefit and harm of pharmacological interventions at the individual patient level is more difficult and less intuitive than in the case of ICDs, but similar scenarios are at play. First, it is true that only some individuals derive benefit and harm and that their numbers drive the signals of benefit or harm observed in clinical trials of pharmacological therapies. Second, the likelihood of benefit in an individual is tightly linked to (1) the likelihood that specific pathogenic mechanisms are present that will both contribute to clinical events in that particular patient and can be modified by the drug (end point-relevant therapeutic targets); (2) the risk of competing causes of adverse outcomes that are not targeted by the drug; and (3) the likelihood of side effects from the drug. Third, a high number needed to treat does not imply that all treated individuals receive a minimally effective intervention but may well signify that many treated individuals do not exhibit abnormalities targeted by the drug or that in many of them, these abnormalities do not lead to clinical events.

To the degree that the processes discussed above can be inferred or predicted by phenotypic, genotypic, and other characteristics, identifying subpopulations of individuals that respond particularly favorably or unfavorably to a drug (the key idea behind precision medicine) should be highly feasible but remains a major challenge. Empirical analyses that detect naturally occurring subgroups of patients who cluster together based on multiple measurable characteristics provide a potentially valuable approach to identifying responders and nonresponders to specific therapies.

In this issue of *Circulation: Heart Failure*, Ferreira et al[3] present an analysis of data from 2 of the 3 largest available trials of mineralocorticoid receptor antagonism (MRA) in HFrEF: the EMPHASIS-HF trial (Eplerenone in Mild Patients Hospitalization and Survival Study in Heart Failure)[4] and the EPHESUS trial (Eplerenone Post-Acute Myocardial Infarction Heart Failure Efficacy and Survival Study).[5] Ferreira et al[3] aimed to identify subgroups of patients with different prognosis and response to eplerenone therapy. The investigators used latent class analysis, a clustering statistical technique that classifies individuals into mutually exclusive and exhaustive subgroups, maximizing within-group similarities and between-group differences on the basis of multiple observed characteristics. The EMPHASIS-HF trial, used for the derivation of clusters (subgroups), enrolled 2737 patients with left ventricular ejection fraction ≤35% and mild (New York Heart Association class II) symptoms and demonstrated that eplerenone produced a 37% reduction in the risk of cardiovascular death and hospitalization in this population.[4] Definitions of clusters derived from the EMPHASIS-HF trial were subsequently applied to the EPHESUS trial, which enrolled 6472 individuals with HFrEF (left ventricular ejection fraction ≤40%) complicating acute myocardial infarction and demonstrated that eplerenone reduced all-cause mortality by 15% in this population.[5] Four subgroups (A–D) of trial participants were identified in EMPHASIS-HF. Participants in subgroup A were more often hypertensive and diabetic. Subgroup B had the highest proportion of lean individuals, a high proportion of participants with anemia and serum K^+<4.0 mmol/L, and a low proportion of participants with glomerular filtration rate <45 mL/min per 1.73 meter^2. Participants in group C were older, often exhibited renal dysfunction, a serum K^+>4.5 mmol/L, previous coronary revascularization, and demonstrated the highest frequency of anemia. Participants in subgroups D and A were younger, had better renal function, less anemia, and higher body mass index; however, subgroup D exhibited a relatively high proportion of participants with K^+>4.5 mmol/L (similar to subgroup C).

Although the subgroups exhibited significant differences in event-free survival, the key merit of the study by Ferreira et al[3] is not survival prediction but the assessment of differential responses to eplerenone (response prediction). Subgroups B and C seemed to obtain a smaller, whereas subgroups A and D seemed to obtain a larger benefit from eplerenone in EMPHASIS-HF (Figure [A]). Subgroups C and D demonstrated the highest rates of eplerenone discontinuation and hyperkalemia during follow-up. The general trend for these responses was also present in EPHESUS (Figure [B]); in the latter trial, subgroup D exhibited a pronounced reduction in the same end point used in EMPHASIS-HF. Interestingly, although groups C and D exhibited the highest rates of hyperkalemia and drug discontinuation during follow-up, subgroup D seemed to enjoy a particularly large benefit from eplerenone. Finally, the 2 subgroups (B and C) that seemed to exhibit a less favorable response in both trials shared common features, such as lower body mass index and a high prevalence of anemia.

The authors provide an online tool that can allocate patients into one of these 4 subgroups using basic phenotypic data. It is important to emphasize the need for its judicious use. This tool should not prevent institu-

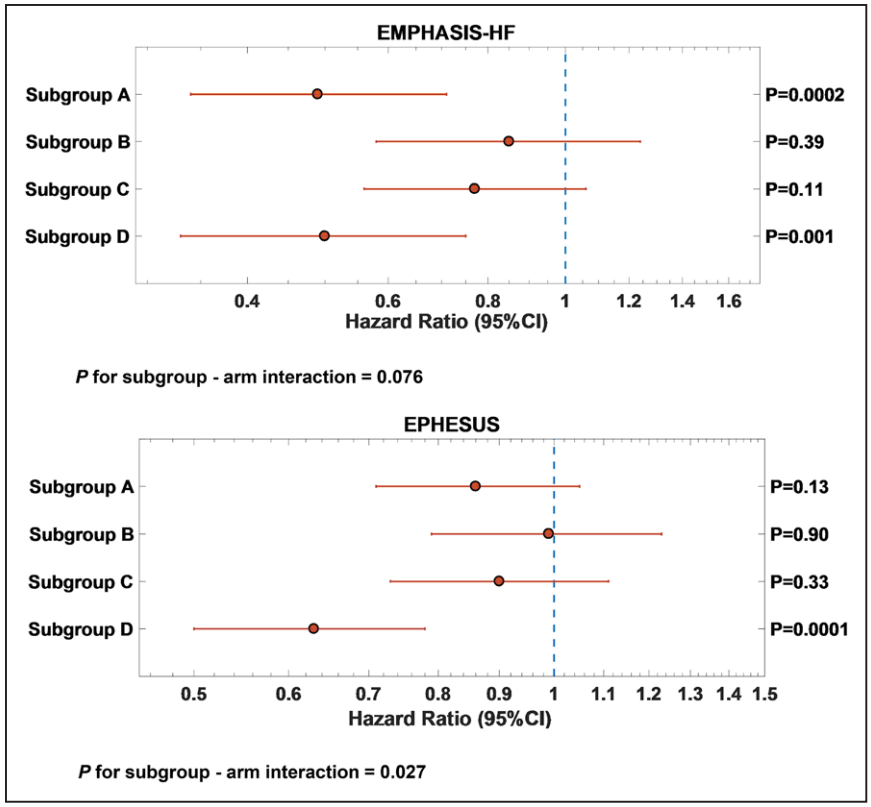

Figure. Hazard ratio for cardiovascular death or heart failure hospitalization in eplerenone vs placebo arm participants in subgroups A–D.
CI indicates confidence interval; EMPHASIS-HF, Eplerenone in Mild Patients Hospitalization and Survival Study in Heart Failure; and EPHESUS, Eplerenone Post-Acute Myocardial Infarction Heart Failure Efficacy and Survival Study. Data from Ferreira et al.[3]

tion of guideline-directed MRA therapy in patients who clearly meet indications. However, the decision to initiate MRA therapy in HFrEF can be complex and impacted by various degrees of renal dysfunction, estimated predisposition to hyperkalemia, the feasibility of monitoring renal function and serum K^+ after initiation of therapy, competing risks that may blunt the potential benefit of MRAs, patient wishes, social issues, etc. The estimation of the risk and benefit of MRA therapy may be enhanced by this tool in various circumstances. In general, the identification of patients in cluster D may more strongly compel the use of MRAs, whereas the identification of patients in clusters C or D may warn the clinician about a higher risk of side effects, prompting closer monitoring.

The study by Ferreira et al[3] provides proof of concept that even basic phenotypic characteristics can help stratify patients in regard to the likelihood of favorable and unfavorable responses to eplerenone. If even basic phenotypic data are indeed informative for such a pleiotropic drug, what could deep phenomics and genomics/pharmacogenomics do for us? Valuable phenotypic data can now be acquired with minimal patient burden in the clinic and can be enhanced by wearable, implanted, and home-based biometric sensors. Could we perform much more accurate assessments of the likelihood of drug responses even for complex, chronic, nonmonogenetic, multifactorial disease states, ultimately improving clinical outcomes? Probably, but these tools will only reach their full potential if they are prospectively engrained in the drug development process and the design of phase III trials. Retrospective analyses of trial data or even prespecified subgroup analyses will generally be insufficient to change practice, and refining the use of already approved therapies is extremely challenging. Consider for instance the hypothetical scenario of a clustering algorithm that, applied retrospectively to the EMPHASIS-HF and EPHESUS trials, identifies a subgroup of subjects for which eplerenone administration was associated with harm. Despite this retrospective evidence, clinicians would still be appropriately compelled to offer eplerenone therapy to all patients who meet overall trial criteria. It would take yet another large phase III trial to overcome such a conundrum, which may never be funded or executed. Adaptive trial designs (such as adaptive enrichment designs)[6] offer an opportunity for incorporating response prediction into phase III or hybrid phase II-III trials but carry their own set of limitations, particularly when most of the enrollment needs to occur upfront, after which a period of observation follows to accumulate events, as often occurs in cardiovascular trials. Reliance on intermediate responses (ie, using biomarkers or other phenotypes) may offer useful information early in the course of a trial, but also introduces potential error and significant uncertainties.

It is worth mentioning that clustering techniques are varied and are rapidly evolving. Latent class analysis is a classic statistical technique that has been used

in social and psychological sciences for decades. More recently, unsupervised machine-learning techniques (such as hierarchical clustering, k-means clustering, gaussian model-based clustering and self-organizing maps) have emerged as powerful techniques for the identification of patient subgroups based on measurable characteristics. It should be noted that different techniques identify different clusters of individuals and even the same technique can provide different results depending on the specification of model parameters and candidate phenotypic characteristics. Therefore, cluster analyses need to be informed by clinical knowledge and judgment, rather than simply by measures of statistical fit or brute computational power. In this regard, carefully chosen phenotypes (particularly ad hoc mechanistic phenotypes) are likely to produce more informative models when it comes to predicting the likelihood of drug response, and these will vary according to the specific drug, end point, and patient population. Expert clinical input will be essential to build reliable algorithms that can better predict drug responses and ultimately enhance clinical practice.

In a not so distant future, heart failure patients may be routinely equipped with various wearable, implanted, or smart home sensors, along with the capability for real-time data transmission and integration. Computational resources capable of accommodating and processing vast amounts of multidimensional data will be available, and some predict that the transfer of authority to algorithms in all realms of life will be a defining feature of this century.[7] How will the phase III trial evolve to exploit the vast opportunities offered by current and upcoming technological innovations? Will classic nonadaptive designs with relatively broad inclusion criteria and minimal mechanistic phenotyping be suitable? The road to precision medicine will be longer and more exciting that we previously thought. We congratulate Ferreira et al[3] for their progressive approach to the analysis of subject-level data to improve our understanding of the effects of aldosterone-receptor antagonists in HFrEF.

ARTICLE INFORMATION

Correspondence

Julio A. Chirinos, MD, PhD, S Tower, Rm. 11–138. Perelman Center for Advanced Medicine, 3400 Civic Center Blvd, Philadelphia, PA 19104. E-mail julio.chirinos@uphs.upenn.edu

Affiliations

Department of Medicine, Cardiovascular Division, Hospital of the University of Pennsylvania and University of Pennsylvania Perelman School of Medicine, Philadelphia (J.A.C.). Department of Medicine, Cardiovascular Division, Henry Ford Heart and Vascular Institute, Henry Ford Health System, Detroit, MI (D.E.L.).

Sources of Funding

This work was supported by R01 HL 121510-01A1 (J.A.C.).

Disclosures

Dr Chirinos has received consulting honoraria from Pfizer, Fukuda Denshi, Ironwood Pharmaceuticals, Sanifit, Bayer OPKO Healthcare and Akros Pharma. He received research grants from National Institutes of Health, American College of Radiology Network, Fukuda Denshi, Bristol Myers Squibb and Microsoft, and device loans from Atcor Medical and Uscom. The other author reports no conflicts.

REFERENCES

1. Bardy GH, Lee KL, Mark DB, Poole JE, Packer DL, Boineau R, Domanski M, Troutman C, Anderson J, Johnson G, McNulty SE, Clapp-Channing N, Davidson-Ray LD, Fraulo ES, Fishbein DP, Luceri RM, Ip JH; Sudden Cardiac Death in Heart Failure Trial (SCD-HeFT) Investigators. Amiodarone or an implantable cardioverter-defibrillator for congestive heart failure. *N Engl J Med*. 2005;352:225–237. doi: 10.1056/NEJMoa043399.
2. Schulz N, Püschel K, Turk EE. Fatal complications of pacemaker and implantable cardioverter-defibrillator implantation: medical malpractice? *Interact Cardiovasc Thorac Surg*. 2009;8:444–448. doi: 10.1510/icvts.2008.189043.
3. Ferreira JP DK, McMurray JJV, Pitt B, van Veldhuisen DJ, Vincent J, Ahmad T, Tromp J, Rossignol P, Zannad F. Data-driven approach to identify subgroups of heart failure with reduced ejection fraction patients with different prognoses and aldosterone antagonist response patterns. *Circ Heart Fail*. 2018;11:e004926. doi: 10.1161/CIRCHEARTFAILURE.118.004926.
4. Zannad F, McMurray JJ, Krum H, van Veldhuisen DJ, Swedberg K, Shi H, Vincent J, Pocock SJ, Pitt B; EMPHASIS-HF Study Group. Eplerenone in patients with systolic heart failure and mild symptoms. *N Engl J Med*. 2011;364:11–21. doi: 10.1056/NEJMoa1009492.
5. Pitt B, Remme W, Zannad F, Neaton J, Martinez F, Roniker B, Bittman R, Hurley S, Kleiman J, Gatlin M; Eplerenone Post-Acute Myocardial Infarction Heart Failure Efficacy and Survival Study Investigators. Eplerenone, a selective aldosterone blocker, in patients with left ventricular dysfunction after myocardial infarction. *N Engl J Med*. 2003;348:1309–1321. doi: 10.1056/NEJMoa030207.
6. Thorlund K, Haggstrom J, Park JJ, Mills EJ. Key design considerations for adaptive clinical trials: a primer for clinicians. *BMJ*. 2018;360:k698.
7. Harari YN. *Homo Deus: A Brief History of Tomorrow*. Harper Collins; 2017.

Circulation
Heart Failure

Embracing the Long Road to Precision Medicine

Julio A. Chirinos and David E. Lanfear

Circ Heart Fail. 2018;11:
doi: 10.1161/CIRCHEARTFAILURE.118.005089
Circulation: Heart Failure is published by the American Heart Association, 7272 Greenville Avenue, Dallas, TX 75231

Print ISSN: 1941-3289. Online ISSN: 1941-3297