Experience With the Reliability of Subgroup Analyses in Trials of Interventions in Heart Failure

Milton Packer, M.D.

Disclosures

- I have spent the last 40+ years leading and analyzing randomized controlled trials in heart failure. I have been the overall Principal Investigator for > 20 trials. I have spent my entire career looking at subgroups.

- I was a member of the Cardiac and Renal Drugs Advisory Committee for 14 years and acted as its chair for 6 years. I retain my current appointment as a Special Government Employee.

- I am representing myself today. Not one is paying for my time. More than a year ago, v-Wave asked for my opinion on their data as a one-time consultation. By my request, my agreement included a clause indicating that I could present my independent views on their data to regulatory agencies. But I will not be expressing any opinion about their data or their device.

Experience With Subgroup Analyses in Heart Failure Trials

- Subgroups are inherently underpowered analyses, based on a baseline variable, to assess the potential influence of the variable on the direction or magnitude of any observed treatment effect (or lack thereof).

- If you perform countless analyses, you will almost always find some subgroup analyses that appear to show an influence of a baseline variable.

- How would you make the judgment that a subgroup finding is real? By ”real”, I mean ”replicable” — you would see the subgroup effect consistently if the trial were carried out again and again.

Some People Believe That It Is Possible to Identify Subgroups That Are *Unlikely* to be Replicated

- Those based on secondary, exploratory or post hoc outcomes
- Those that are based on few events
- Those that are difficult to explain clinically.
- Those that are based on a variable that is associated with exceptional variability in its assessment (e.g., ejection fraction)

Some People Believe That It Is Possible to Identify Subgroups That Are Likely to be Replicated

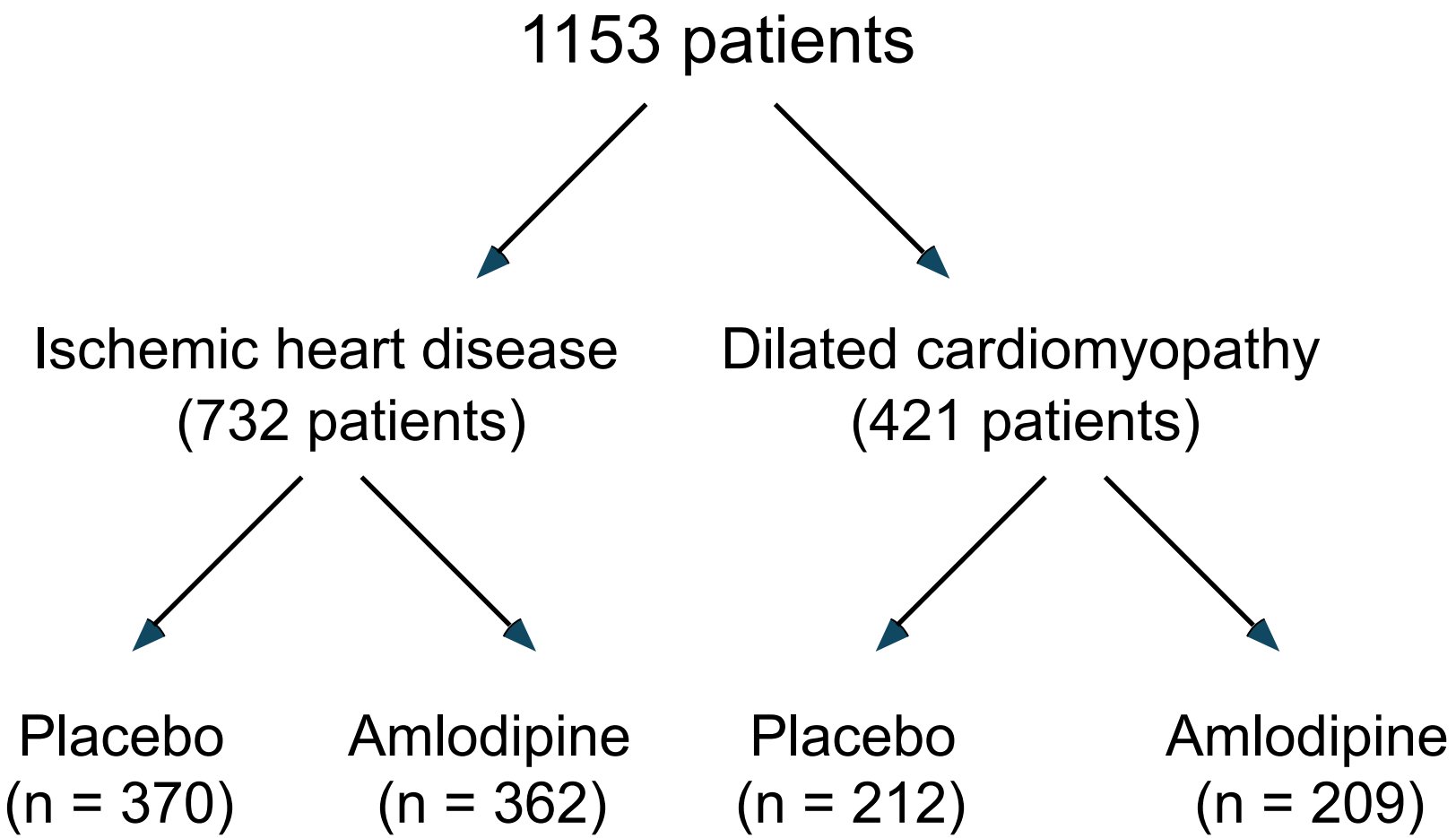
- Those based on the primary endpoint.
- Those that are prespecified.
- Those based on a stratification variable.
- Those that are biologically plausible.
- Those that have a statistically significant interaction P value.

Some People Believe That It Is Possible to Identify Subgroups That Are Likely to be Replicated

- Those based on the primary endpoint. **NO!**
- Those that are prespecified. **NO!**
- Those based on a stratification variable. **NO!**
- Those that are biologically plausible. **NO!**
- Those that have a statistically significant interaction P value. **NO!**

Almost always, “subgroup interactions” cannot be replicated

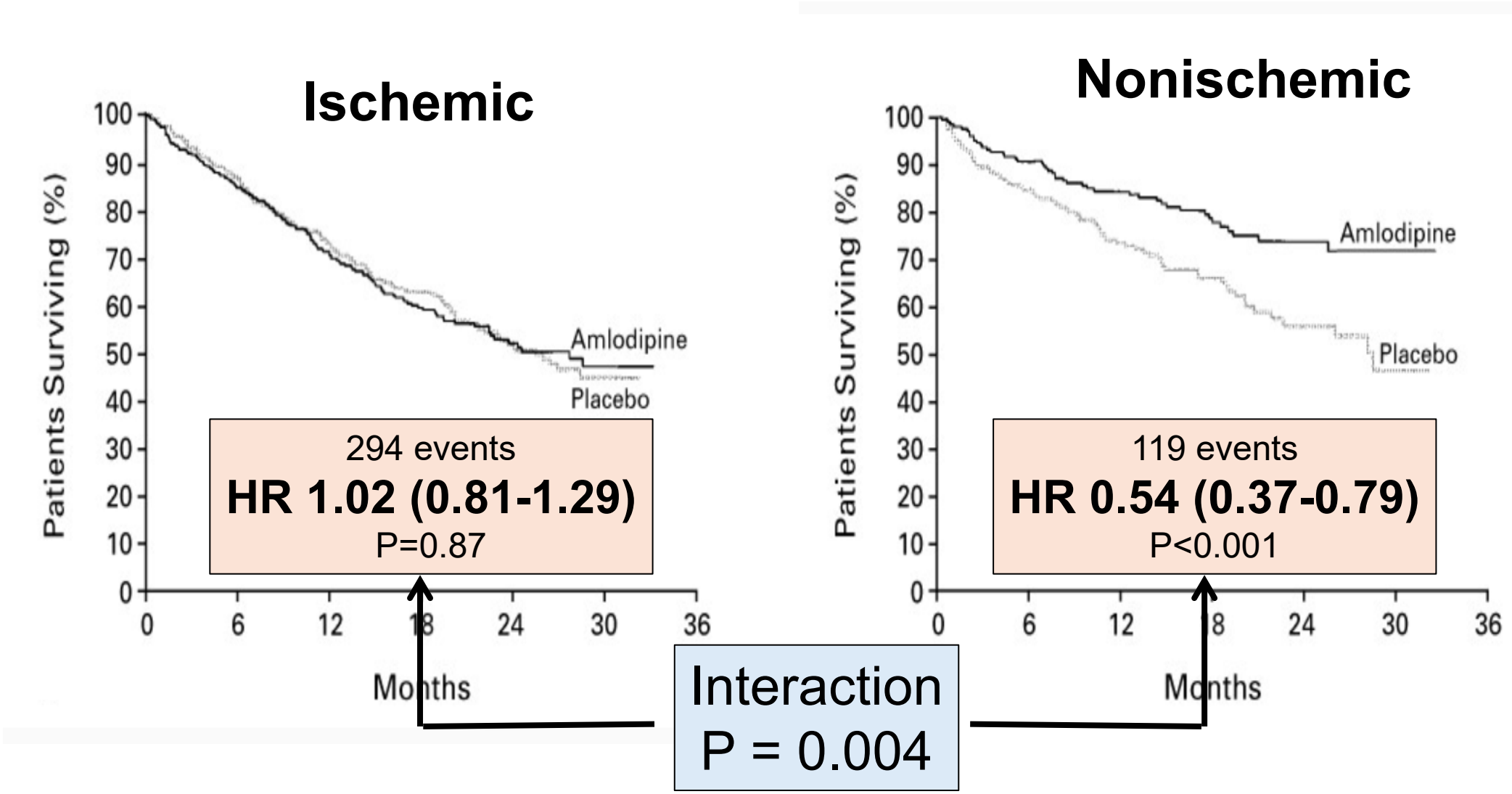
PRAISE-1: Enrollment, Stratification and Randomization



PRAISE-1: Effect on All-Cause Mortality in All Patients (Ischemic and Nonischemic Combined)

	Placebo	Amlodipine	Hazard Ratio	P-Value
All patients	223/582 (38.3%)	190/571 (33.3%)	0.84 (0.69-1.02)	0.070

PRAISE-1: Effect of Amlodipine vs Placebo on All-Cause Mortality in Subgroups



Did We Have a Mechanism to Explain This? Of Course, We Did!

Physicians can always think of a mechanism. **Always**.

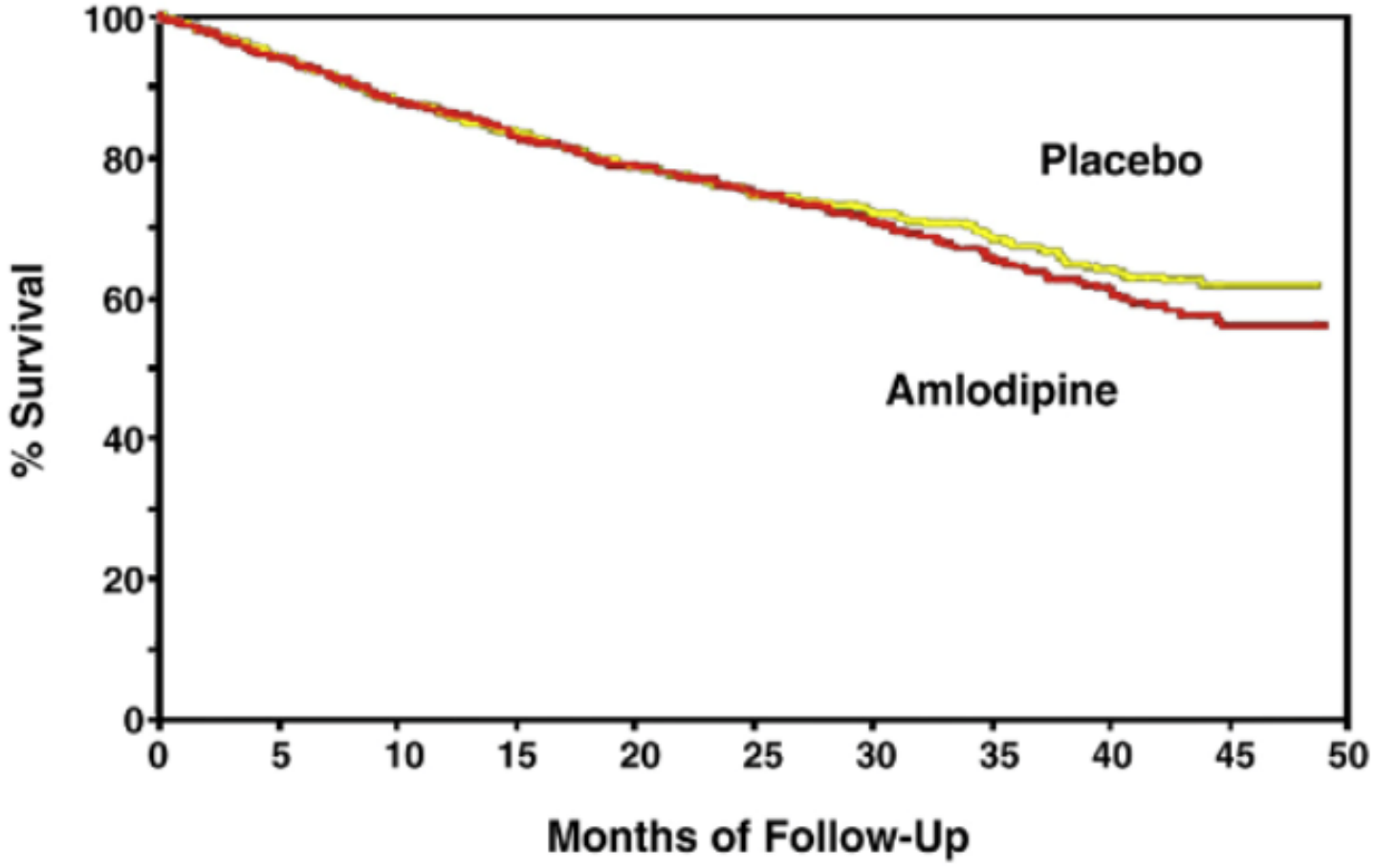
- We thought patients with nonischemic cardiomyopathy were likely to have coronary microvascular vasospasm, as in Takotsubo cardiomyopathy. Amlodipine would block this mechanism.
- We had several substudies on different variables.
- But imaging or biomarker substudies performed ***in the same trial*** are problematic, since any play of chance that causes an apparent effect on "events" in any subgroup would "cluster" with surrogates. In the same trial, patients who do well have surrogates that do well. But they do not inform the likelihood of replication.

To Know If the Effect Was Real, We Needed to Do Another Trial. So PRAISE-2 Was Launched to Replicate the Striking Subgroup Finding in PRAISE-1

The two trials (PRAISE-1 and PRAISE-2)

- Used the same protocol, with same definition of nonischemic cardiomyopathy and utilized the same dose of amlodipine.
- Relied on the original group of investigators
- Recruited patients with same baseline characteristics and receiving the same background therapy

PRAISE-2: Effect on All-Cause Mortality in Nonischemic Cardiomyopathy



PRAISE-2 had longer follow-up with a larger number of events

HR 1.09 (0.92, 1.29)
P=0.32

NO REPLICATION of the subgroup finding in nonischemic cardiomyopathy in PRAISE-1

HR 0.54 (0.37-0.79)
P< 0.001

Conclusions

- Even under conditions of prespecification and stratification and even when highly significant interaction P values are seen on endpoints of exceptional clinical relevance, treatment subgroup effects seen in trials of interventions in patients with chronic heart failure are rarely replicated.

- Therefore, replication of a subgroup finding is essential to know if any observed subgroup effect is actually real.

- We particularly worry about the actionability of subgroups showing a qualitative interaction, where one group seems to benefit and the complementary subgroup seems to be harmed — especially when the decision is determined by a variable that is not measured with precision in clinical practice.

Thank you.

I am happy to take any questions

that members of the Committee might have.