Heros

Lem Moyé, MD, PhD

Lem Moyé
UTHealth School of Public Health – Houston
1200 Herman Pressler E1009
Houston, TX 77030
LemMoye@msn.com

Cardiology researchers must look to themselves to save their research programs. Biostatisticians, however innocently have placed the cardiovascular research community in danger and we must oursides decide on a different course.

Recently, the results of the VEST trial were presented. The findings of no impact of a VEST to reduce the primary endpoint of sudden cardiac death/arrhythmic death for which there was no statistically significant difference in event rates (1.6% vs 2.4%, $p = 0.18$), but a positive finding for a secondary endpoint, has produced confusion as to what the results mean. Many wish to believe the mortality findings but are cautioned against that because it was not primary.

Cardiology researchers have been here before. There are strict rules for interpretation of clinical research that have not been so helpful. The use of statistical hypothesis testing and *p*-values in the Multiple Risk Factor Intervention Trial (MRFIT) [1], followed by the experiences of the vesnarinone program [2-3], the ELITE trials [4-5], PRAISE [6,7] and the Carvedilol program [8, 9, 10, 11, 12, 13, 14] generated a collection of interpretative conundrums confronting the cardiology research community.

These results are confusing and play a role in the reproducibility crisis in health care research. The response by some is to call for smaller p-values. Yet this produces a greater focus on the intrepretation of a smaller number of endpoints, aggravating, and not alleviating the problem.

Is it not time for cardiologists to take responsibility for the interpretion of the clinical trials in its field of expertise and require new rules for interpretation?

The process of selecting and relying on a small number of analyses to represent the findings of clinical trial is analysis parsimony. It is a bedrock principal of clinical trials.

It is time to ask what is the reason for this parsimony? Its basis is not epidemiologic. Epidemiologists argue that multiple analyses should be conducted and reported. One justification is to encourage an assessment of internal validity. A second is to compare findings across studies, requiring that variables measured in other studies also be measured in the current trial. Nor can the motivation for analysis frugality be found in the bedrock fields of chemistry, biochemistry, physiology or pathology. There is nothing stated or implied in the scientific method that constrains the investigator from collecting and reporting supporting data and analyses.

The principal reason for analysis parsimony is rooted in how sampling error is managed. We biostatisticians customarily conduct a formal statistical hypothesis on each endpoint, accumulating type I error as we proceed across the hypotheses . This accumulation is precise, but is not what investigators need or find helpful. Investigators nevertheless acquiesce in this process, ultimately finding themselves stymied from answering the question they designed the study to address because the statistical analysis was only designed to evaluate the smaller questions dichotomously, not the larger ensemble question.

The path that clinical trialists have tread to this destination is well documented [15,16]. However, given that the twin threats of crippling cost and excessive inefficiency threaten to strangle clinical research, it is time to ask anew whether clinical trial methodology is strengthened or weakened by the use of statistical hypothesis testing?

A helpful perspective is gained by viewing the clinical trial as an evidence gathering tool, collecting information to either support or refute a research question. The purpose of the investigator is to collect and then in conjunction with the research community weigh the evidence. For this process to be productive, the evidence must have two properties.

Property 1: Evidence must be relevant. The relevance of clinical trial evidence is wholly within the purview of the investigators. The investigators choose the research question, then choose the data and analyses that are logically linked to the research question at hand. Investigators by their nature and the cost effectiveness of data collection are commonly driven to collect a wealth of information to help them determine the best answer to the question. These answers can be found simultaneously at multiple levels (vital status, hospitalization rates, morbidity data, imaging information, laboratory data, and biomarker information). Investigators actively engage in the collection of this data because of their beliefs, based on both science and experience, that this information is helpful in assessing the individual's response to the exposure being tested in the trial and therefore is connected to the research question.

Property 2: Evidence must be reliable. Clinical investigators commonly think of reliability as a measure of instrument precision. We biostatisticians have expanded this notion of reliability to include the concept of sampling error. The recognition that sample findings do not always reflect population results is critical. However, there are clear differences between the reactions of biostatisticians and the responses of clinical researchers to this observation. The reaction of the biostatistician is to require that each analysis assessment generate a statistically based decision and to pay a type I error penalty for that decision. This approach forces the investigator to decide for each analysis, whether that analysis confirms or refutes a hypothesis.

However, clinical researchers never intended to assess each piece of evidence as either wholly supportive or wholly unsupportive of the global hypothesis. They instead desire to measure each piece of evidence for the degree to which it supports or does not support a scientific idea. To clinical scientists, the issue is more subtle than a 0-1 statistical litmus test. A change in blood pressure, for example, given its natural variability can be supportive of the possibility of benefit and the possibility of harm simultaneously. Clinical investigators understand that biology is nuanced. Thus from the perspective of the clinical investigator every conducted analysis based on precisely measured data makes a contribution to the overall finding. Each is a continuum and each is probative.

The consequences of the biostatistical approach versus the investigator approach to evidence reliability in clinical trials cannot be ignored. Because we biostatisticians require type I error control, the universe of "reliable" evidence from a clinical trial is much smaller than that of all relevant evidence. The notion of one or two primary endpoints, buttressed by a small number of secondary endpoints is a direct ramification of type I error control. All other evidence is inadmissible to the biostatistician since the overall family-wise error rate would be exceeded. Thus to the biostatistician, a small number of analyses is dispositive for the research question.

However, the continuum philosophy embraced by investigators suggests that each relevant analysis provides information about the effect of the exposure and therefore each must be considered. By factoring in the variability associated with the analysis, investigators understand that the result is not dichotomous, but instead demonstrates duality; the analysis finding can simultaneously support the concept of both benefit and harm. It is only through the accumulation of these many pieces of evidence, each suggesting possible affirmation and possible refutation of the hypothesis, that researchers can reach a reliable conclusion.

The impact of these two approaches also affects result synthesis. According to the statistical decision theory perspective, very little integration of trial results is needed. If there is one primary endpoint, then *ceteris paribus* that is all that is required; the one analysis is decisive. However, the continuum approach requires substantial integration. There are many well-designed, well-conducted analyses in clinical trials. While some provide evidence that is wholly affirmative, or wholly dismissive of the scientific question at hand, most provide evidence to support both. How can this information, reporting different degrees of support or rejection of the scientific question, whose effect sizes are measured in different units, be helpfully and quantitatively integrated into a fair assessment of the conclusion drawn from all of the evidence?

Thus clinical trials should be analysed to address the following questions:

1- What evidence is there in the trial to address the salient research issue "Are participants who are exposed to the intervention better off than those exposed to the control?" and what is the strength of that evidence.
2- What is the strength of evidence supporting benefit? What is the estimate of that benefit?
3- What is the estimate supporting harmt? What is the estimate of that harm?

Unfortunately, there are no such tools. However, these implements of evidence integration are desperately required by clinical trialists who wish to improve the efficiency of the expensive clinical trials that the public health community holds in such high regard and also answer the difficult global questions for which they were designed. While work has begun on this, more work and workers are needed.

Fellow cardiologists – no one is riding to your rescue. You must become your own heros and ask for what you need.

References

1 . Multiple risk factor intervention trial. Risk factor changes and mortality results. Multiple Risk Factor Intervention Trial Research Group. JAMA. 1982 Sep 24;248(12):1465-77.

2. Feldman AM, Bristow MR, Parmley WW et.al (1993) Effects of vesnarinone on morbidity and mortality in patients with heart failure. *N Engl J Med* 329:149-55

3. Cohn J, Goldstein SC, Feenheed S et.al. (1998) A dose dependent increase in mortality seen with vesnarinone among patients with severe heart failure. *N Eng J Med* 339:1810-16

4. Pitt B, Segal R, Martinez FA. et al. on behalf of the ELITE Study Investigators (1997). Randomized trial of losartan versus captopril in patients over 65 with heart failure. *Lancet* 349:747–52.

5. Pitt B, Poole-Wilson PA., Segal R, et. al (2000). Effect of losartan compared with captopril on mortality in patients with symptomatic heart failure randomized trial–The losartan heart failure survival study. ELITE II. *Lancet*.355:1582–87.

6. Packer, M., O'Connor, C.M., Ghali, J.K., et al for the Prospective Randomized Amlodipine Survival Evaluation Study Group (1996). Effect of amlodipine on morbidity and mortality in severe chronic heart failure. *New England Journal of Medicine.*335:1107–14.

7. Packer, M. (2000). Presentation of the results of the Prospective Randomized Amlodipine Survival Evaluation-2 Trial (PRAISE-2) at the American College of Cardiology Scientific Sessions, Anaheim, CA, March 15, 2000.

8. Packer M., Bristow MR. Cohn JN. et al. (1996). The effect of carvedilol on morbidity and mortality in patients with chronic heart failure. *New England Journal of Medicine.* 334:1349-55.

9. Moyé LA, Abernethy D. (1996) Carvedilol in Patients with Chronic Heart Failure (Letter). *New England Journal of Medicine.* 335: 1318-1319.

10. Packer M, Cohn JN., Ccolucci WS. Response to Moyé and Abernethy. *New England Journal of Medicine* 335:1318-1319.

11. Fisher L. (1999). Carvedilol and the FDA approval process: the FDA paradigm and reflections upon hypothses testing. *Controlled Clinical Trials* 20:16-39.

12. Fisher LD, Moyé LA. (1999) Carvedilol and the Food and Drug Administration Approval Process: An Introduction. *Controlled Clinical Trials*. 20:1-15.

13. Moyé L.A. (1999) *P* Value Interpretation in Clinical Trials. The Case for Discipline. *Controlled Clinical Trials* 20:40-49.

14. Fisher LD. Carvedilol and the Food and Drug Administration-Approval Process: A Brief Response to Professor Moyé's article. *Controlled Clinical Trials*. 20:50-51.

15 Goodman, S.N. (1999). Toward Evidence–Based Medical Statistics. 1: The p-value fallacy. Annals of Internal Medicine 130:995–1004.

16 . Moyé LA (2003) *Multiple Analyses in Clinical Trials: Fundamentals for Investigators* New York. Springer. Prologue