

# Can we live without $P$ values? The answer



Eugene H. Blackstone, MD

From the Department of Thoracic and Cardiovascular Surgery, Heart and Vascular Institute, and the Department of Quantitative Health Sciences, Research Institute, Cleveland Clinic, Cleveland, Ohio.

Disclosures: Author has nothing to disclose with regard to commercial support.

Received for publication Sept 7, 2017; accepted for publication Sept 8, 2017; available ahead of print Oct 6, 2017.

Address for correspondence: Eugene H. Blackstone, MD, Department of Thoracic and Cardiovascular Surgery,

Cleveland Clinic, 9500 Euclid Ave, Desk JJ-4, Cleveland, OH 44195 (E-mail: [blackse@ccf.org](mailto:blackse@ccf.org)).

*J Thorac Cardiovasc Surg* 2018;155:1137

0022-5223/\$36.00

Copyright © 2017 by The American Association for Thoracic Surgery

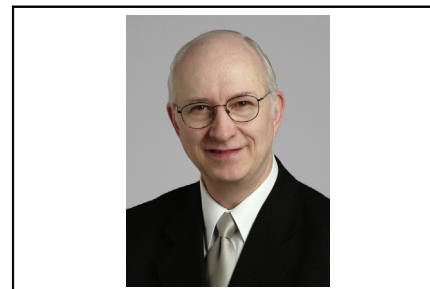
<https://doi.org/10.1016/j.jtcvs.2017.09.030>

When members of the American Statistical Association's board decided to develop a policy statement on  $P$  values and statistical significance, they did so knowing they had not previously taken positions on specific matters of statistical practice.<sup>1</sup> They recognized that "misunderstanding or misuse of statistical inference" had reached a point that a policy statement was necessary: "In view of the prevalent misuses of and misconceptions concerning  $P$ -values, some statisticians prefer to supplement or even replace  $P$ -values with other approaches." The common use of  $P$  values to identify risk factors in multivariable models, which Naftel<sup>2</sup> called more of an art than science, sprang to mind. Could one really "replace  $P$  values" and actually better answer research questions? I challenged Hemant Ishwaran, PhD, one of our Deputy Statistical Editors, to consider how one might identify "significant" risk factors in commonly used multivariable models without using  $P$  values.

Applying methods borrowed from machine learning, he and University of Miami graduate student Min Lu developed a novel method to accomplish this. In their article in this issue of the *Journal*, Lu and Ishwaran<sup>3</sup> introduce two measures of "variable importance" as a substitute for  $P$  values and illustrate them by reanalyzing heart failure data using well-known Cox proportional hazards regression.

The method is more than a substitute for  $P$  values, however. Built in is bootstrap resampling, whereby new data sets are formed and analyzed by sampling the original set with replacement. This means that some observations are repeated and others are left out—about a third on average—allowing statistics to be generated based on how well results are predicted for the left-out patients. "Important" variables are those that improve this prediction. I cannot overemphasize the huge advantage that this provides. We often chide authors who claim to have "predictive models" but provide no internal (let alone external) validation to be able to use the word "predictor." The Lu-Ishwaran method provides thousands of cross-validations as a byproduct.

Another advantage is less sensitivity to sample size than  $P$  values. This is important for analyses of genomic data and large national data sets in which extremely tiny  $P$  values may be produced for nearly every variable, so that it



Eugene H. Blackstone, MD

### Central Message

Methods borrowed from machine learning can be used to identify important variables in ordinary multivariable analyses without use of  $P$  values.

See Article page 1130.

becomes unclear which are the important features. Further, their method is a gateway into a host of other machine-learning methods, such as efficient ways to manage large numbers of variables and ways to illuminate the shape of relationships of continuous variables to outcomes without model assumptions.

There are limitations, however. The method is computationally intensive, but those who already routinely borrow machine-learning methods to generate multivariable models are used to this.<sup>4</sup> It does not provide statistics with which we are familiar. Fortunately, Efron and Hastie<sup>5</sup> from Stanford have recently published an accessible book intended to bridge statistics of the past, including  $P$  values, and those of the computer age.

Can we live without  $P$  values? Perhaps not for research (such as clinical trials) well suited to a method modeled on English common law (innocent until proven guilty beyond reasonable doubt). But for variable selection using common multivariable models, it may well be possible to live, and live well, without  $P$  values!

### References

1. Wasserstein RL, Lazar NA. The ASA's statement on  $P$ -values: context, process, and purpose. *Am Stat*. 2016;70:129-33.
2. Naftel DC. Do different investigators sometimes produce different multivariable equations from the same data? *J Thorac Cardiovasc Surg*. 1994;107:1528-9.
3. Lu M, Ishwaran H. A prediction-based alternative to  $P$  values in regression models. *J Thorac Cardiovasc Surg*. 2018;155:1130-6.
4. Rajeswaran J, Blackstone EH. Identifying risk factors: challenges of separating signal from noise. *J Thorac Cardiovasc Surg*. 2017;153:1136-8.
5. Efron B, Hastie T. *Computer age statistical inference: algorithms, evidence, and data science*. New York: Cambridge University Press; 2016.