To the Editor:

We read with interest the article by Taylor et al. This article highlights the potential of random forest to correctly classify sepsis in-hospital mortality. The authors demonstrate superior predictive performance of random forest over methods traditionally used in emergency medicine, classification and regression tree (CART) and a generalized linear mixed model (GLMM), by comparing the area under the ROC curves.

Random forest is an improvement over CART because it averages over many (bootstrap aggregated) trees, so it was not a surprise that it outperformed CART, which fits only one tree. Although random forest tends to fall under the title of machine learning, it certainly does not encompass the set of candidate algorithms that might have also worked well at predicting death in these patients. Perhaps a better representation of applying machine learning to these data would have included five- or 10-fold cross-validation (CV) to compare the mean-squared-error of a greater number of algorithms. Additional algorithms to consider in this data set would include adaptive least absolute shrinkage and selection operator, because the data were clustered. Additionally, gradient boosting machine is an algorithm that will often outperform random forest because it adaptively improves with each fit of a tree. Finally, when comparing the predictive validity of models, it seems unfair to not first determine the single best parsimonious linear model. Although the authors performed preselection of 20 variables from the many available in their data set based on literature review, they did not do any variable selection on the GLMM fit to their data, and they made no mention of testing for nonlinearity or interactions. Perhaps the GLMM would have performed better if they had done this step first. It is well known that if a model is overfit to the training data it will have higher variability in the test set.

This study is an important demonstration of the predictive power of modern machine learning techniques, applied to large clinical data sets captured through the electronic health record. To fully harness the power of machine learning, algorithms in addition to random forest should be tested and compared using k-fold CV. Diagnosis and prediction are vital to the clinical and research mission of emergency medicine. We agree with the authors that machine learning methods are techniques to include more often in emergency medicine research.

doi: 10.1111/acem.13009

Halden Scott, MD
(halden.scott@childrenscolorado.org)
Departments of Pediatrics and Emergency Medicine
School of Medicine University of Colorado Denver
Children’s Hospital Colorado

Kathryn Colborn, PhD
Division of Health Care Policy and Research
Adult and Child Consortium for Health Outcomes Research and Delivery Science
School of Medicine University of Colorado Denver
Denver, CO

Supervising Editor: John H. Burton, MD.

References