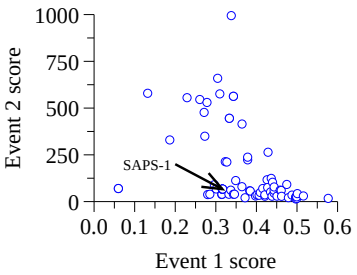
Predicting Mortality of Patients in Intensive Care: The PhysioNet/Computing in Cardiology Challenge 2012

Ikaro Silva, George B Moody, Daniel J Scott, Leo A Celi, Roger G Mark

Massachusetts Institute of Technology
Cambridge, MA, USA

Acuity scores such as APACHE, SAPS, MPM, and SOFA are widely used to account for population differences in studies aiming to compare how medications, care guidelines, surgery, and other interventions impact mortality in Intensive Care Unit (ICU) patients. By contrast, the focus of the PhysioNet/CinC Challenge 2012 is to develop methods for patient-specific prediction of in-hospital mortality.

The data used for the challenge consist of up to 41 general descriptors (age, gender, height, weight) and time series (hourly measurements of vital signs and laboratory test results) from the first 48 hours of the first available ICU stay of each of 12,000 patients chosen at random from a larger set. Patients under age 16 and those whose initial ICU stays were shorter than 48 hours (approximately the median) were excluded; there were no other exclusion criteria. We randomly divided the data into three sets (A, B, and C) of 4,000 patients each. Challenge participants were provided data and outcomes for set A, and they submitted algorithms to estimate risk and predict survival or death for individual subjects. In Phase 1, submitted algorithms were tested using set B, which participants were allowed to study although outcomes were withheld. Each entry was scored in event 1 according to its utility for prediction (the lower of sensitivity and positive predictivity), and in event 2 according to the accuracy and utility of its risk estimates (using a range-normalized Hosmer-Lemeshow statistic). A baseline algorithm (SAPS-1) scored 0.316 in event 1 and 66.6 in event 2. Most participants submitted Phase 1 entries that outperformed the baseline algorithm; scores were as high as 0.577 in event 1 and as low as 13.9 in event 2. Participants may continue to refine their methods during Phase 2, which culminates in a blinded test of the most successful methods using set C.



Results of 88 Phase 1 entries from 39 participants (best at lower right).