Statistical pitfalls and errors in external validation studies of clinical prediction models: a case study with the SAPS II and SAPS 3 prognostic models for critically ill patients

Evangelos I. Kritsotakis [1], Stamatios Kokkinakis [2], Ioannis Xanthakis [3], Augusto Morgia [4], Maria Samaritaki [5], Konstantinos Lasithiotakis [2]

[1] *Laboratory of Biostatistics, School of Medicine, University of Crete, Heraklion, Greece*

[2] *Department of General Surgery, School of Medicine, University of Crete, Heraklion, Greece.*

[3] *Master of Biostatistics Program, Medical School and Department of Mathematics, National Kapodistrian University of Athens, Athens, Greece*

[4] *Department of Public Health and Infectious Diseases, Specialization School of Health Statistics and Biometrics, Sapienza Università Di Roma, Roma, Italy*

[5] *Department of Anaesthesiology, University Hospital of Heraklion, Heraklion, Greece*

INTRODUCTION: Successful application of a clinical prediction model requires external validation by applying the model as originally specified to different patient populations or settings and then quantifying the accuracy of the predictions made. Empirical accounts examining how researchers apply statistical methods for this purpose are lacking. Herein we report statistical pitfalls and errors identified from a systematic review of the prognostic performance of the Simplified Acute Physiology Score (SAPS) II and SAPS 3 in intensive care units (ICUs). The SAPS models rely on ICU admission data to estimate the probability of subsequent in-hospital death with a logistic regression equation.

METHODS: PubMed and Scopus were searched for studies externally validating the discrimination and calibration performances of SAPS II or 3 in ICU patients, since 2003. Forward citation searches were performed and previous systematic reviews of SAPS II/3 were surveyed for eligible studies. Data were extracted using the CHARMS checklist. Risk of bias (RoB) was assessed with the PROBAST tool. PROSPERO CRD42023446721.

RESULTS: In all, 89 external validation reports involving 495,443 patients in 1,157 ICUs in 33 countries were assessed. Most studies were performed in Europe (n=38; 43%), Asia (n=26; 29%), or South America (n=13; 15%). Most involved prospective data collection (n=57; 64.0%), sampled consecutive ICU admissions (n=84; 94%), and were designed either as cohort (n=69; 78%) or registry-based (n=20; 22%) studies. In the PROBAST analysis domain, 73 reports (82%) were classified as high RoB and 8 (9%) were unclear RoB, whereas only 8 reports (9%) were low RoB. Most concerns were raised due to inappropriate handling of missing data by either unjustified complete case analysis (n=43; 48%) or single imputation by normal/reference values (n=12; 13%). Other concerns were inadequate evaluation of model performance measures (n=30; 34%), low numbers of outcome events (n=29; 33%), and not including all enrolled patients in the analysis (n=20; 23%). Most studies did not report a formal sample size calculation (n=85; 96%). Calibration was assessed solely with a statistical test in 13 cases (15%). Discrimination was examined with the AUC in all studies, but 20 (22%) calculated a classification measure after dichotomizing SAPS with some cut-off threshold. Decision curve analysis was performed in only 3 studies (3.4%).

CONCLUSION: External validation studies of prognostic models are viewed as necessary and are pursued by the medical research community. However, this investigation reveals an urgent need for improved methodological frameworks, reporting transparency and statistical literacy when validating risk models.

KEYWORDS: Prediction model, risk of bias, methodological quality, validation, systematic review.

TOPIC: Prediction and Prognostic Models