Journal of Clinical Epidemiology 78 (2016) 83−89

Modern modeling techniques had limited external validity in predicting mortality from traumatic brain injury

Tjeerd van der Ploeg[a,b,c,*], Daan Nieboer[c], Ewout W. Steyerberg[c]

[a]Department of Science, Medical Center Alkmaar, Wilhelminalaan 12, Alkmaar 1815 JD, The Netherlands
[b]Department of Science, Inholland University, Bergerweg 200, Alkmaar 1817 MN, The Netherlands
[c]Department of Public Health, Erasmus MC—University Medical Center Rotterdam, P.O. Box 2040, 3000 CA Rotterdam, The Netherlands

Accepted 5 March 2016; Published online 14 March 2016

Abstract

Background and Objective: Prediction of medical outcomes may potentially benefit from using modern statistical modeling techniques. We aimed to externally validate modeling strategies for prediction of 6-month mortality of patients suffering from traumatic brain injury (TBI) with predictor sets of increasing complexity.

Methods: We analyzed individual patient data from 15 different studies including 11,026 TBI patients. We consecutively considered a core set of predictors (age, motor score, and pupillary reactivity), an extended set with computed tomography scan characteristics, and a further extension with two laboratory measurements (glucose and hemoglobin). With each of these sets, we predicted 6-month mortality using default settings with five statistical modeling techniques: logistic regression (LR), classification and regression trees, random forests (RFs), support vector machines (SVM) and neural nets. For external validation, a model developed on one of the 15 data sets was applied to each of the 14 remaining sets. This process was repeated 15 times for a total of 630 validations. The area under the receiver operating characteristic curve (AUC) was used to assess the discriminative ability of the models.

Results: For the most complex predictor set, the LR models performed best (median validated AUC value, 0.757), followed by RF and support vector machine models (median validated AUC value, 0.735 and 0.732, respectively). With each predictor set, the classification and regression trees models showed poor performance (median validated AUC value, <0.7). The variability in performance across the studies was smallest for the RF- and LR-based models (inter quartile range for validated AUC values from 0.07 to 0.10).

Conclusion: In the area of predicting mortality from TBI, nonlinear and nonadditive effects are not pronounced enough to make modern prediction methods beneficial.

Keywords: Prediction models; Modeling techniques; Internal validation; External validation; Discrimination; Calibration

1. Introduction

Prediction of binary outcomes has since long received much attention in medical research. Prediction is complicated by the specification of the model structure, such as the inclusion of main effects, potential nonlinearities, and statistical interactions [1−3]. Although most prediction models for binary end points are still based on logistic regression (LR) analysis, there is increasing interest in other, more modern techniques, such as support vector machines (SVMs), neural nets (NNs), and random forests (RFs). These more modern methods hold the promise of better capturing nonlinearities and interactions in medical data [4].

A decisive factor in choosing a modeling technique for prediction is the performance of the resulting model at external validation. Many studies compared modern modeling techniques with classical techniques, but mostly, they only validated the resulting models internally [5,6]. External validation was used in only a few comparisons of classification trees, neural networks and LR [7,8], and in a comparative study on stroke patients [9].

In this study, we aimed to compare the external validity of LR and four more modern modeling techniques to predict 6-month mortality of patients suffering from traumatic

Conflict of interest: The authors wish to confirm that there are no conflicts of interest associated with this publication, and there has been no financial support for this work that could have influenced its outcome. We confirm that the manuscript has been read and approved by all authors and that there are no other persons who satisfied the criteria for authorship but are not listed. We further confirm that the order of authors listed in the article has been approved by all of us.

* Corresponding author. Tel.: 0031-653519264.
E-mail address: tvdploeg@quicknet.nl (T. van der Ploeg).

http://dx.doi.org/10.1016/j.jclinepi.2016.03.002
0895-4356/

What is new?

Key findings

- Despite the availability of modern prediction techniques, classical logistic regression may still be optimal for prediction in new patients in a low-dimensional setting with adequate sample size such as mortality of traumatic brain injury.

What this adds to what was known?

- It is the first time that modern techniques, such as support vector machines, neural nets, and random forests were externally validated for predicting 6-month mortality in TBI patients.

What is the implication and what should change now?

- Since performance may vary substantially across settings, external validation is a necessary step before applying a prediction model in a new setting, specifically if a model was developed with a relatively modern technique.

brain injury (TBI). We chose this patient group because TBI is a heterogeneous disease, in which many mechanisms and pathways can lead to mortality and poor long-term outcome [10−13]. Moreover, tree-based models have specifically been suggested to be beneficial for prediction of outcome after TBI [4]. In patients with moderate or severe injuries, mortality 6 months after surgery exceeds 20% and lifelong disability occurs in half of the survivors [14]. Prediction of outcome in patients with TBI using prediction models has been studied since the 1970s [15,16]. However, the preferred technique for prediction of outcome of TBI patients is still under debate, and preference for a technique varies between investigators [4]. Various statistical techniques have been used in this area, including LR, recursive partitioning, Bayesian approaches, and discriminant analysis [16]. Nowadays, a wide array of modern learning techniques is available, including RFs, SVMs, and neural networks [1,17]. We investigated whether nonlinear and nonadditive effects in the area of predicting mortality from TBI are pronounced enough such that these modern modeling techniques can outperform traditional modeling techniques such as LR.

2. Methods

2.1. Patients

We analyzed individual patient data from the IMPACT database [14,18,19]. This database includes data of patients suffering from moderate or severe TBI. The database comprises data from 11,026 patients included in 15 different studies (Appendix A, Table 1, Fig. 1/Appendix at www.jclinepi.com). Patients were enrolled in one of ten randomized clinical trials or in one of five registries between 1984 and 2006.

2.2. Modeling techniques

We compared five statistical modeling techniques to predict 6-month mortality:

- Logistic regression (LR)
- Classification and regression trees (CART)
- Random forests (RFs)
- Support vector machines (SVMs)
- Neural nets (NNs)

We here list the main characteristics of the evaluated modeling techniques, based on previous work of several authors [2,3,17,20,21]. We refer to Appendix 3/Appendix C at www.jclinepi.com for the code of our analyses in R software (Vienna, Austria) [22].

2.2.1. Logistic regression

LR is a type of regression analysis that is often used in medical research to model the probability of a binary end point using a linear function of the predictors. Predictor variables may be either continuous or categorical. LR uses a logistic transformation to calculate the probability of a binary outcome. Regression coefficients were estimated by maximum likelihood using the *lrm* function in the *rms* library.

2.2.2. Classification and regression trees

CART is modeling technique that uses recursive partitioning to split the training records into segments with similar end point values. The modeling starts by examining the input variables to find the best split, measured by the reduction in an impurity index that results from the split. The split defines two subgroups, each of which is subsequently split into two further subgroups and so on, until a stopping criterion is met. The commonly used parameter for CART is the cp-parameter (cost complexity factor). A cp-value of 0.001 for example regulates that a split must decrease the overall lack of fit by a factor of 0.001. The modeling was done using the *rpart* function in the *rpart* library.

2.2.3. Random forest

RF is an ensemble classifier that consists of many decision trees. In case of classification, RF outputs the class that is the mode among the classes from individual trees. In case of regression, RF outputs the value that is the mean of the values output from individual trees. Each tree is constructed using a bootstrap sample from the original data. A tree is grown by recursively partitioning the bootstrap sample based on optimization of a split rule. In regression