Noname manuscript No.
(will be inserted by the editor)

A Novel Pediatric Mortality Risk Prediction Score Based on Nonlinear Feature Transformations

Natalia M. Arzeno · Karla A. Lawson · Sarah V. Duzinski · Haris Vikalo

Received: date / Accepted: date

Abstract The PRISM III score is widely used in pediatric intensive care units (PICU) and has been extensively validated in various settings. While simple enough to allow for fast manual evaluation, PRISM III dichotomizes predictive variables to form prediction scores, which may lead to critical failures. In this paper, we seek to develop a risk prediction score that preserves clinical knowledge embedded in features and structure of PRISM III while addressing limitations caused by variable dichotomization. The novel method transforms predictive variables using nonlinear logistic functions that allow for a fine differentiation between critical and normal values of the variables. Optimal parameters of the logistic functions are inferred for a given patient population via cyclic block coordinate descent, and may be readily re-learned as patient population and standards of care evolve. We tested the proposed technique on brain trauma patients admitted to the PICU of the Dell Children's Medical Center of Central Texas between 2007 and 2012. The prediction power of the score is evaluated using area under ROC curve (AUC), Youden's index J, and precision-recall balance in a leave-one-out cross-validation study. The results

This material is based upon work supported by the National Science Foundation Graduate Research Fellowship under Grant No. DGE-1110007.

N.M. Arzeno
Department of Electrical and Computer Engineering, The University of Texas at Austin, Austin, TX, 78712 USA
E-mail: narzeno@utexas.edu

K.A. Lawson and Sarah V. Duzinski
Trauma Services, Dell Children's Medical Center of Central Texas
4900 Mueller Blvd Austin, TX 78723

H. Vikalo
Department of Electrical and Computer Engineering, The University of Texas at Austin
1 University Station C0803, Austin, TX 78712
Tel.: +1-512-232-7922 E-mail: hvikalo@ece.utexas.edu

demonstrate that the new score significantly outperforms PRISM III in terms of all three criteria.

Keywords Prognosis · Trauma · Brain injury · Optimizable risk score · Continuous risk score

1 Introduction

Technological advancements in medical instrumentation and a growing use of electronic medical records have created an abundance of clinical patient data. Extracting and analyzing useful information from such large and diverse data sets will enable tremendous advancements in clinical decision-making, ultimately leading towards improvements in health and quality of life as well as to reduction of the overall healthcare costs. Availability of data has enabled development of accurate mortality and morbidity risk prediction scores for specific patient populations. Rapid prediction of potentially poor outcomes may provide timely intervention to reduce morbidity and mortality among the patients in the considered group.

The Pediatric Risk of Mortality (PRISM III) (Pollack et al, 1996) is a widely used mortality risk prediction score in pediatric intensive care units (PICU) (Marcin and Pollack, 2000) that has been validated in various settings as both an individual predictor and a significant predictor in a multivariate model in the United States and internationally (Scavarda et al, 2010; Gemke and van Vught, 2002; Brady et al, 2006; Karambelkar et al, 2012; Cantais et al, 2001; Bahloul et al, 2011; Volakli et al, 2012; Martha et al, 2005; Qureshi et al, 2007). However, while simple enough to allow for fast manual evaluation, PRISM III dichotomizes predictive variables to form prediction scores. Dichotomization of continuous variables results in a loss of information, increased probability of false negatives, and high dependence on cut-off points (Streiner, 2002; Royston et al, 2006). This, in turn, may lead to critical failures of the prediction process. Moreover, note that PRISM III is a relatively mature prediction mechanism that was developed in 1996. To remain relevant, prediction scores need to be updated periodically so that they reflect innovations and the evolution of the standards of healthcare (Marcin and Pollack, 2000).

In this paper, we seek to develop a risk prediction score that preserves clinical knowledge embedded in features and structure of PRISM III while allowing for soft thresholds in the score calculation. The paper is organized as follows. Some of the most common risk scores used in the PICU are reviewed in Section 2. In Section 3, we describe an algorithm for finding optimal values of the logistic functions used to transform the data and discuss data imputation strategies. Section 4 presents results obtained after applying the proposed methodology to predicting outcome of pediatric trauma patients admitted to the PICU of the Dell Children's Medical Center of Central Texas between 2007 and 2012. Section 5 concludes the paper.

2 Related Work

A state-of-the-art outcome prediction scheme, PRISM III, has an additive step-wise structure that relies on 17 physiological variables and 26 ranges. The physiological variables it considers are characterized by their maximum or minimum values recorded during the first 12 or 24 hours after a patient's admission to the PICU. These variables include the minimum systolic blood pressure, maximum heart rate, the presence of fixed pupils, maximum and minimum body temperature, and a variety of laboratory measures including minimum and maximum CO_2 and pH, white blood cell count, glucose and platelet count. Contribution of some variables to the score is determined after evaluating a judiciously chosen logical OR statement, such as in the case of maximum prothrombin time (PT) and partial thromboplastin time (PTT), which can both detect abnormalities in clotting time. Other variables, such as systolic blood pressure and heart rate, have age-dependent ranges.

In PRISM III, the score is incremented when the maximum (minimum) value of a variable in the score is above (below) a predetermined threshold. For example, if a child has a minimum Glasgow Coma Scale (GCS) (Laurer et al, 2002; Maas et al, 2011) score less than 8, then 5 points are added to her/his PRISM III score. Clearly, calculation of the score is highly dependent on the established cutoff points and thus the prediction may change abruptly due to very small changes in the underlying variables. For instance, interpretation of the variables such as heart rate or blood pressure, which are altered by the simple act of breathing, may widely change due to the strict threshold structure – an adolescent with a maximum heart rate of 144 beats per minute is considered healthy, while another one with a single measure of 145 beats per minute has 3 extra points added in the PRISM III score calculation. Although the PRISM scores have been validated in numerous settings, they have also been shown to overpredict (Slater et al, 2004; Tibby et al, 2002) and underpredict (Bhadoria and Bhagwat, 2008; Thukral et al, 2006) PICU deaths. Poor patient discrimination by PRISM scores, especially in neonates and infants (Wells et al, 1996; Goddard, 1992), and the fact that only a small subset of PRISM variables are significant predictors of the outcome (Ponce-Ponce De León et al, 2005), render the PRISM scores sensitive to population characteristics and standards of care and suggest that they may not necessarily be institution independent. In this paper we develop a novel score that can be optimized for a specific patient population. Moreover, we go beyond simply finding new coefficients for the features in the existing scoring mechanisms and instead develop a novel methodology that identifies a range of values for which the risk changes continuously and monotonically (i.e., increases or decreases with the feature values). This is accomplished by characterizing the effect of a feature onto the risk by a non-linear logistic curve whose inflection point and slope we find via an optimization, as opposed to using hard thresholding to describe the effect of the physiological variables on the risk of mortality which is the procedure employed by PRISM III and other state-of-the-art prediction schemes.

In addition to PRISM III, widely-used scores in the PICU include the injury severity score (ISS) (Baker et al, 1974) and the pediatric index of mortality (PIM2) (Slater et al, 2003), where the former is specific to trauma patients. The ISS is an anatomic score based on the location and severity of the injuries. Limitations of the ISS have led to various modifications as well as risk scores that incorporate the ISS in the calculation (Marcin and Pollack, 2002). In a pediatric trauma population, PRISM has outperformed the ISS and its variants in identifying in-hospital mortality (Cantais et al, 2001). Logistic regression has often been employed to learn the weighting coefficients for physiologic variables or binary indicators in mortality prediction models (Marcin and Pollack, 2002). PIM2 is a second generation score, based on recalibrating coefficients of PIM (Shann et al, 1997) and adding variables for diagnostic groups with poor performance or calibration. PIM2 models risk using logistic regression with 10 variables acquired upon hospital admission or in the first hour after PICU admission, 7 of which are binary indicators. Specifically, the continuous PIM2 variables include first systolic blood pressure, ratio of FiO_2 to PaO_2, and absolute arterial or capillary base excess. Fixed pupils, mechanical ventilation, elective admission, PICU admission for procedure recovery, and cardiac bypass are included in the model as binary variables. Finally, a selection of high risk or low risk diagnoses, where the model was found to over- or under-estimate mortality, complete the list of logistic regression variables (Slater et al, 2003).

Methods recently used for the development of risk prediction scores include decision tree techniques (Courville et al, 2009), autoregressive implementation of PRISM (Ruttimann and Pollack, 1993), and techniques that incorporate injury coding schemes in the models (Burd and Madigan, 2009). Most pediatric mortality risk prediction scores assume that risk depends linearly on the variables and do not consider nonlinear variable transformations. PRISM III is an exception since it uses thresholding to prevent uninformative increase in risk at very high or very low variable values. In neonate morbidity prediction, the PhysiScore (Saria et al, 2010) achieves greater predictive power than previously established neonatal morbidity scoring systems by relying on a nonlinear transformation of the raw variables in the feature set. In particular, that work employs nonlinear Bayesian models based on log odds ratios of the risk derived from the probability distribution that provides the best fit to the data for each of two patient classes.

Models that integrate domain expert knowledge with a data driven approach have been reported to result in greater predictive accuracy. In (Sun and Hu, 2012), combining knowledge based and data driven risk factors in a prediction model for heart failure greatly improved on the performance of a solely knowledge based classifier while still resulting in a clinically meaningful model. In the task of identifying similar concept pairs in clinical notes, combining context-based similarity and knowledge-based similarity in an algorithm has likewise resulted in a more accurate similarity score (Pivovarov and Elhadad, 2012).

3 Methods

In this section we propose a novel outcome prediction score, present an algorithm for optimizing parameters of the function that transforms predictive features, and discuss optimality of the aforementioned algorithm.

3.1 The new score and an algorithm for optimizing parameters of the logistic transformation of predictive features

We describe the risk of mortality using a logistic regression model, where the conditional probability that patient i dies during the hospital stay is given by

$$\mathbf{P}(y_i = 1|\mathbf{w}, \mathbf{z}_i) = \frac{1}{1 + \exp(-\mathbf{w}^T \mathbf{z}_i)}, \tag{1}$$

and the conditional probability of survival is

$$\mathbf{P}(y_i = -1|\mathbf{w}, \mathbf{z}_i) = \frac{1}{1 + \exp(\mathbf{w}^T \mathbf{z}_i)}, \tag{2}$$

where $y_i \in \{-1, 1\}$ is an indicator of the in-hospital mortality, the vector $\mathbf{w} = [w_1\ w_2\ \dots w_{M_w}]'$ collects weights for the features $\mathbf{z}_i = [z_{i1}\ z_{i2}\ \dots\ z_{iM_w}]'$, and M_w denotes the total number of features.

In a departure from the commonly used hard thresholding of predictive features and discrete scoring (as is done in PRISM III), we introduce a logistic transformation of the predictive features. The resulting new score is continuous and differentiable which enables computationally efficient search for the optimal parameters of the logistic transformation. In particular, for patient i and feature j, the nonlinear transformation z_{ij} of the raw variable x_{ij} is

$$z_{ij} = \begin{cases} \frac{1}{1+\exp(-a_j(x_{ij}-t_{ij}))} & \text{if } x_{ij} \text{ is a maximum} \\ 1 - \frac{1}{1+\exp(-a_j(x_{ij}-t_{ij}))} & \text{if } x_{ij} \text{ is a minimum} \\ 0 & \text{if } x_{ij} \text{ is missing} \end{cases} \tag{3}$$

where $a_j \geq 0$ is the slope of the nonlinear transformation and t_{ij} is the inflection point of the logistic function (i.e., a "soft threshold" counterpart to the hard thresholds used by the existing scoring schemes such as PRISM). It should be noted that t_{ij}'s are not different for every subject i but rather have the same value for all patients within a specific age group and a given feature j if such age group dependence is suggested by the existing clinical knowledge.

The optimal weights and parameters for the nonlinear transformations are determined by minimizing the negative log-likelihood of the logistic regression model,

$$\min \sum_{i=1}^{n} \log(1 + \exp(-y_i(\mathbf{w}^T \mathbf{z}_i))). \tag{4}$$

To preserve and exploit clinical knowledge previously used in the creation of other scores, a lognormal prior is imposed in the optimization for $\mathbf{w}$; this also ensures all features will be associated with a positive weight. In particular, for $\mathbf{w} \in \mathbb{R}^d$ we set

$$\mathbf{P}(\mathbf{w}) = \frac{\exp\left(-\frac{1}{2}(\log \mathbf{w} - \mu)^T \Sigma^{-1} (\log \mathbf{w} - \mu)\right)}{(2\pi)^{d/2} |\Sigma|^{0.5} \prod_{j=1}^{d} w_j}.$$

For a lognormal prior with mean μ and covariance $\Sigma = \frac{1}{2\lambda} I$, the optimization over $\mathbf{w}$ then becomes

$$\begin{aligned} \arg\min_{\mathbf{w}} & \sum_{i=1}^{n} \log(1 + \exp(-y_i(\mathbf{w}^T \mathbf{z}_i))) \\ & + \sum_{j=1}^{d} \log w_j + \lambda \| \log \mathbf{w} - \mu \|_2^2. \end{aligned} \tag{5}$$

The joint optimization (5) over $\mathbf{w}$, $\boldsymbol{a} = [a_1 \ a_2 \ \dots \ a_{M_a}]$, and/or $\mathbf{t} = [t_1 \ t_2 \ \dots \ t_{M_t}]$ is carried out by cyclic block coordinate descent with backtracking line search (Boyd and Vandenberghe, 2009). Optimization over $\boldsymbol{a}$ and $\mathbf{t}$ includes an additional step of projections onto the constraint set. The blocks for the coordinate descent consist of features derived from the same raw variable. For example, one block contains two features derived from the maximum heart rate, which correspond to two steps with different weights in a simple thresholding-based additive score. The algorithm is formalized as Algorithm 1 given below. Note that the objective function of the optimization (5) is not convex. Nevertheless, even if we use an iterative optimization method that merely ensures the objective function is decreased at each step, the resulting local minimum leads to an improvement over PRISM III. In fact, our computational studies show that the proposed scheme is robust with respect to the initial point of the search – starting the iterative optimization procedure with a vector $\mathbf{w}$ comprising PRISM III weights and starting with a vector of uniform weights $\mathbf{w} = ([1, \cdots, 1]$ results in almost identical prediction accuracy on the dataset of interest.

In Algorithm 1, G is the set of feature groups, A denotes the projection set for $\boldsymbol{a}$, and α and β are the backtracking line search parameters. The feature groups $g \in G$ are defined as the groups of nonlinear features related to the same raw variables (e.g., the feature group for maximum pH has two elements corresponding to PRISM III thresholds of 7.55 and 7.48), and make up the blocks for the coordinate descent.

The projection sets are defined so that the clinical knowledge used for PRISM III is preserved. In Algorithm 1, A ensures the nonlinear transformations follow the same direction as the steps in PRISM III ($A = \{\boldsymbol{a} : a_j \geq 0$ for $j = 1, \dots, M_a\}$). The projection set T for optimization over the soft thresholds of the nonlinear transformations in (3), $\mathbf{t}$, preserves the order of soft thresholds for a raw variable with multiple nonlinear transformations. For example,

Algorithm 1 Optimization over the slopes $\boldsymbol{a}$

$a_j^{(0)} \leftarrow 0.01,\ j = 1, 2, ..., M_a$
$\mathbf{w} \leftarrow \mathbf{w}_{\text{PRISM}}$
$\mathbf{t} \leftarrow \mathbf{t}_{\text{PRISM}}$
$k \leftarrow 1$
repeat
  $\boldsymbol{a} \leftarrow \boldsymbol{a}^{(k-1)}$
  for all $g \in G$ **do**
    $\triangle a_j \leftarrow -(\nabla_a f(\boldsymbol{a}, \mathbf{t}, \mathbf{w}))_j$ if $j \in g$
    $\triangle a_j \leftarrow 0$ if $j \notin g$
    $h \leftarrow 1$
    while $f(\boldsymbol{a} + h\triangle\boldsymbol{a}, \mathbf{t}, \mathbf{w}) > f(\boldsymbol{a}, \mathbf{t}, \mathbf{w}) - \alpha h||\triangle\boldsymbol{a}||^2$ **do**
      $h \leftarrow \beta h$
    end while
    $\boldsymbol{a} \leftarrow \text{Proj}_A(\boldsymbol{a} + h\triangle\boldsymbol{a})$
  end for
  $\boldsymbol{a}^{(k)} \leftarrow \boldsymbol{a}$
until stopping criterion is met

if feature i in PRISM III corresponds to the maximum heart rate between 215 and 225 beats per minute ($t_i^{(0)} = 215$) and feature j corresponds to the maximum heart rate above 225 beats per minute ($t_j^{(0)} = 225$), then the projection ensures that $t_i \leq t_j$ at all steps of the optimization procedure.

The dimensions of the three optimization parameters are not equal since binary features associated with pupillary reaction do not have nonlinear transformations ($M_a < M_w$) and some of the thresholds are age-dependent ($M_w < M_t$). Optimization over the slopes $\boldsymbol{a}$ and soft thresholds $\mathbf{t}$ are implemented without inclusion of a prior in the objective.

3.2 Quasiconvexity of the logistic transformation

As stated earlier in this section, the objective function in (5) is not convex. However, we will here show that each block in the block-coordinate descent procedure is both quasiconvex and quasiconcave in the slope parameter $\boldsymbol{a}$, and is thus quasilinear. Since the logistic function is asymptotically flat, the objective in (5) is not strictly quasiconvex. However, since the objectives in the steps of block-coordinate descent procedure are quasilinear, if the initial values of $\boldsymbol{a}$ are such that the gradient is nonzero in every coordinate, the block coordinate descent will reach a global optimum.

For differentiable f and domain $\mathcal{D}$, f is quasiconvex if and only if $\mathcal{D}$ is convex and for all $x, y \in \mathcal{D}$ holds that $f(y) \leq f(x) \Rightarrow \nabla f(x)^T(y - x) \leq 0$ (Boyd and Vandenberghe, 2009). Similarly, f is quasiconcave if and only if $\mathcal{D}$ is convex and for all $x, y \in \mathcal{D}$ it holds that $f(y) \geq f(x) \Rightarrow \nabla f(x)^T(y - x) \geq 0$.

The objective of the optimization is

$$\min f = \sum_{i=1}^{n} \log(1 + \exp(-y_i(\mathbf{w}^T\mathbf{z}_i))), \tag{6}$$

where

$$\mathbf{w}^T\mathbf{z}_i = \sum_{j\in U} \frac{w_j}{1+\exp(-a_j(x_{ij}-t_{ij}))} + \sum_{k\in D} w_k\left(1-\frac{1}{1+\exp(-a_k(x_{ik}-t_{ik}))}\right) + \sum_{p\in P} w_p\delta(p_i=1), \tag{7}$$

where U denotes the set of indices of features with maximum values whose contribution to the score is in the form of an up-step (i.e., the risk is higher when their values are above a threshold), D is the set of indices of features with minimum values (down-steps), P is the set of indices of the pupillary reflex features taking binary values $\{0,1\}$, and $\delta(p_i=1)$ is an indicator function for the pupillary reflex features. The pupillary reflex features indicate whether one or both pupils are >3mm and fixed.

Note that if we want to find when is $f(a'_j) \le f(a''_j)$ for a given a'_j and a''_j, it is sufficient to find the conditions on $a'_j, a''_j, y_i, x_{ij}, t_{ij}$ such that it holds that $f_i(a'_j) \le f_i(a''_j)$ for all $i \in 1,\dots n$, where $f = \sum_{i=1}^n f_i$. For any $j \in U$, condition $f(a'_j) \le f(a''_j)$ is satisfied on the domain where

$$\frac{-y_i w_j}{1+\exp(-a'_j(x_{ij}-t_{ij}))} \le \frac{-y_i w_j}{1+\exp(-a''_j(x_{ij}-t_{ij}))}.$$

This inequality will hold for any of the parameter combinations marked by an X in Table 1. The blank spaces in Table 1 satisfy $f_i(a'_j) \ge f_i(a''_j)$ for $j \in U$,

Table 1 Satisfy $f_i(a'_j) \le f_i(a''_j)$ for $j \in U$

	$a'_j \le a''_j$		$a'_j \ge a''_j$	
	$y_i = 1$	$y_i = -1$	$y_i = 1$	$y_i = -1$
$x_{ij} - t_{ij} \ge 0$		X	X	
$x_{ij} - t_{ij} \le 0$	X			X

the fact which we next use to show quasiconcavity. To show quasilinearity, we also need to examine the gradient of f at a''_j which, for $j \in U$, is given by

$$\nabla_{a_j} f_i(a''_j) = -\frac{\exp(-y_i(\mathbf{w}^T\mathbf{z}_i))}{1+\exp(-y_i(\mathbf{w}^T\mathbf{z}_i))} \times \frac{y_i w_j (x_{ij}-t_{ij})\exp(-a''_j(x_{ij}-t_{ij}))}{\left(1+\exp(-a''_j(x_{ij}-t_{ij}))\right)^2}. \tag{8}$$

The sign of the gradient in (8) is determined by $-y_i w_j(x_{ij}-t_{ij})$. Simple arithmetic shows that when $f_i(a'_j) \le f_i(a''_j)$ (corresponding to the entries marked by X in Table 1), $\nabla_{a_j} f_i(a''_j)(a'_j - a''_j) \le 0$, and hence the condition for

quasiconvexity is satisfied. Similarly, when $f_i(a'_j) \geq f_i(a''_j)$ (corresponding to blanks in Table 1), $\nabla_{a_j} f_i(a''_j)(a'_j - a''_j) \geq 0$, which implies quasiconcavity.

The same procedure can be followed to show quasilinearity when $j \in D$, with appropriate sign changes. Note that the blocks of features where all of the features in the block belong to U or all of the features in the block belong to D will also satisfy the quasilinearity condition. Since the coordinate blocks in Algorithm 1 correspond to nonlinear transformations of the same physiologic variable, quasilinearity holds. The initial value of a_j in Algorithm 1 is set to 0.01 for all j. This corresponds to a small slope in the nonlinear transformations of the variables, and will only result in $\nabla_{a_j} f_i(a_j) = 0$ for subjects i with $x_{ij} = t_{ij}$. Therefore, the initial slope of the cumulative function in (6) will be nonzero and the coordinate descent algorithm will not begin at a stationary point.

Since we constrain the slopes $\boldsymbol{a}$ and weights $\mathbf{w}$ to be nonnegative, quasilinearity of the objective function in soft thresholds $\mathbf{t}$ can be shown in fewer steps than that for $\boldsymbol{a}$. We will only show quasiconvexity in t_{ij} when $j \in U$, but quasiconvexity when $j \in D$ and quasiconcavity can be easily shown in a similar manner. To demonstrate quasiconvexity, it is sufficient to show that

$$f_i(t'_{ij}) \leq f_i(t''_{ij}) \Rightarrow \nabla_{t_{ij}} f_i(t''_{ij})(t'_{ij} - t''_{ij}) \leq 0.$$

Note that for $j \in U$, $f_i(t'_{ij}) \leq f_i(t''_{ij})$ if $y_i = -1$ and $t'_{ij} \geq t''_{ij}$, or if $y_i = 1$ and $t'_{ij} \leq t''_{ij}$. By examining the gradient of f_i, it is clear that since $w_j, a_j \geq 0$, $\nabla_{t_{ij}} f_i(t''_{ij}) \leq 0$ only when $y_i = -1$,

$$\begin{aligned}\nabla_{t_{ij}} f_i(t''_{ij}) = &\frac{\exp(-y_i(\mathbf{w}^T \mathbf{z}_i))}{1 + \exp(-y_i(\mathbf{w}^T \mathbf{z}_i))} \times \\ &\frac{y_i w_j a_j \exp(-a'_j(x_{ij} - t''_{ij}))}{\left(1 + \exp(-a_j(x_{ij} - t''_{ij}))\right)^2}.\end{aligned} \tag{9}$$

Note that the required conditions for $f_i(t'_{ij}) \leq f_i(t''_{ij})$ also guarantee that $\nabla_{t_{ij}} f_i(t''_{ij})(t'_{ij} - t''_{ij}) \leq 0$. However, unlike with the optimization over slopes $\boldsymbol{a}$, we cannot guarantee that the initial $\mathbf{t}$ will be at a point with nonzero slope. Nevertheless, we have empirically observed that although $\nabla_{t_{ij}} f_i = 0$ for some (i, j), $\sum_i \nabla_{t_{ij}} f_i \neq 0$ in the first iteration of the alternating optimization of $\boldsymbol{a}$ and $\mathbf{t}$, indicating the optimization does not start at a stationary point for any of the nonlinear features. As with the weights $\mathbf{w}$, we initialize $\mathbf{t}^{(0)}$ using the PRISM III thresholds. Despite the approximations made in order to simplify calculation of the PRISM III score, one expects its thresholds to be close to the optimal values due to the reliance on domain expert knowledge and extensive testing. Therefore, the proposed initialization will likely avoid start of the optimization procedure in the flat part of the quasi convex curve, implying the global optimality of the block coordinate descent.

3.3 Algorithm Testing

Data were retrospectively abstracted for 217 children (11.06% mortality rate) admitted to the Dell Children's Medical Center PICU. We included admissions to the PICU between August 2007 and April 2012, age range of 0-14 years, with an ICD-9 code reflecting brain injury, and a PICU stay of at least 24 hours. The ICD-9 codes indicate patients with brain trauma (excluding simple concussion) as well as other brain malady or injury such as cerebral palsy, drowning, epilepsy and asphyxiation. This group of patients was selected in order to emphasize the ease of optimizing the novel score for a specific high mortality population. Minimum and maximum variable values from the first 12 hours of the PICU stay are used in the calculation of PRISM III and our new score. The use of the data has been approved by the University of Texas at Austin Institutional Review Board and the Seton Clinical Research Steering Committee.

The proposed prediction scheme is tested using leave-one-out cross-validation and compared with existing methods in terms of three criteria: (i) area under the receiver operating characteristic (ROC) curve (AUC), (ii) the Youden index (J), which aims to maximize the overall correct classification rate (Hilden and Glasziou, 1996; Perkins and Schisterman, 2006): $J = \text{Sensitivity}(Se) + \text{Specificity}(Sp) - 1$, and (iii) the point on the ROC curve that maximizes the minimum of the positive predictivity ($+P$, precision) and sensitivity (recall). The last criterion takes into account class imbalance by balancing the percentage of true positives that are correctly predicted with the percentage of predicted positives that are correct. Thus, correctly predicted true negatives, the majority class, does not affect the performance metric. It should be noted that in evaluating algorithms by the third criterion (PrecRec), a false positive has the same effect as a false negative.

4 Results

4.1 Implementation of the novel prediction scheme

Our novel score incorporates all of the variables and ranges previously used by PRISM III. The features transformed using OR statements in PRISM III (e.g., by adding 6 points to the score if pH < 7.0 OR total $CO_2 < 5$) were treated as additive features in order to seamlessly include them in the optimization procedure. The backtracking line search parameters were chosen empirically based on the accuracy and speed of convergence. The results shown are for $\alpha = 0.2$ and $\beta = 0.5$. In the case of optimization over $\mathbf{w}$, the imposed prior has $\mu = 0, \lambda = 0.25$.

Table 2 Mortality and Missing Data by Age Group

Age Group	No. Subjects	No. Deaths	Missing Variables (mean ± SD)
Neonate (0 mo., 1 mo.)	2	0	9 ± 11.31
Infant [1 mo., 12 mo.)	39	5	6.95 ± 7.13
Child [12 mo., 144 mo.]	143	15	7.02 ± 7.22
Adolescent (144 mo., 180 mo.)	33	4	6.03 ± 7.14

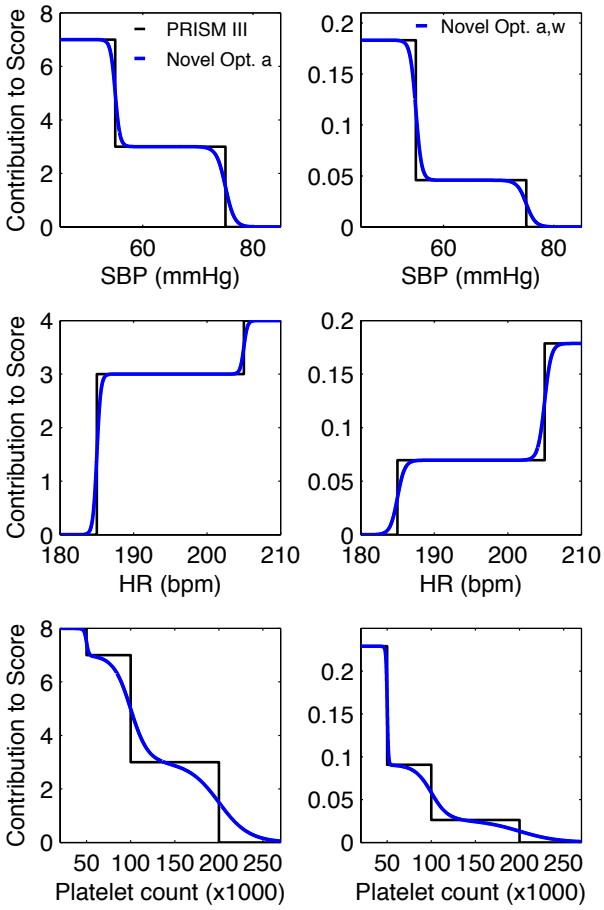

Fig. 1 Contribution of systolic blood pressure (SBP), heart rate (HR), and platelet count to risk score for PRISM III (black, left) and novel score optimizing $\boldsymbol{a}$ (blue, left) and optimizing $\boldsymbol{a}, \mathbf{w}$ (blue,right). The black lines on the right panel serve as indicators of a sharp transition between steps.

4.2 Performance comparison

Figure 1 illustrates the difference in computing the contribution of a feature to the prediction score between the 12-hour PRISM III and the novel scheme.

Table 3 Risk Score Accuracy

Score	AUC	J	PrecRec
PRISM III	0.8735	0.5840	0.5172
Novel Score optimized over $\boldsymbol{a}$	0.8897	0.6215	0.5000
Novel Score optimized over $\mathbf{w}$	0.8841	0.6369	0.5600
Novel Score optimized over $\boldsymbol{a}$ and $\mathbf{t}$	0.8358	0.6153	0.5833
Novel Score optimized over $\boldsymbol{a}$ and $\mathbf{w}$	0.8927	0.6682	0.5833
PIM 2	0.8331	0.6729	0.5417

Plots on the left show the logistic transformations of the features after performing optimization over the nonlinear transformation slopes $\boldsymbol{a}$, while the plots on the right show the logistic transformations after performing optimization over both the slopes $\boldsymbol{a}$ and weights $\mathbf{w}$. From the plots, we see that the optimization over $\boldsymbol{a}$ results in mortality risk increasing over a range of 3 to 4 beats per minute for each step in the maximum heart rate and the risk from minimum systolic blood pressure increasing over a range of 7 mmHg around the 75 mmHg threshold and increasing over a range of 4 mmHg around the 55 mmHg threshold. The risk from the minimum platelet count loses much of the stepwise scoring structure used by PRISM III, and instead increases monotonically for minimum platelet counts between 250,000 and 50,000. This illustrates how our novel score can capture risk that increases continuously throughout a certain range while maintaining sharp thresholds when those are optimal. Optimizing over both $\boldsymbol{a}$ and $\mathbf{w}$ provides further insight into the contribution of variables to the risk of mortality for a given population. For example, the novel score shows that, for this dataset, the second step in the nonlinear transformation of the systolic blood pressure and heart rate should be weighted similarly, if not more heavily, than the lower step. The lower panels of Figure 1 also indicate that while PRISM III has the mortality risk increasing slightly more when the minimum platelet count falls below 50,000, our inferred scoring function indicates the risk contribution from this variable doubles at approximately 50,000.

The results of the ROC analysis on the cross-validated scores are shown in Table 3. Our novel score optimized over the slopes $\boldsymbol{a}$ of the nonlinear transformations results in a more accurate classifier than PRISM III, in terms of the AUC and J, but not in terms of the precision-recall balance (PrecRec). Optimization of the feature weights $\mathbf{w}$ results in a score that performs better than PRISM III in all three evaluation criteria. Alternating optimization over the slopes $\boldsymbol{a}$ and weights $\mathbf{w}$ results in a further improvement of AUC, J and precision-recall balance over individual parameter optimization, and thus provides a significant advancement over PRISM III. The inclusion of soft thresholds $\mathbf{t}$ along with slopes $\boldsymbol{a}$ in the alternating optimization results in some of the soft thresholds falling outside the physiological range of the raw variables – in particular, logistic transformations of these variables result in

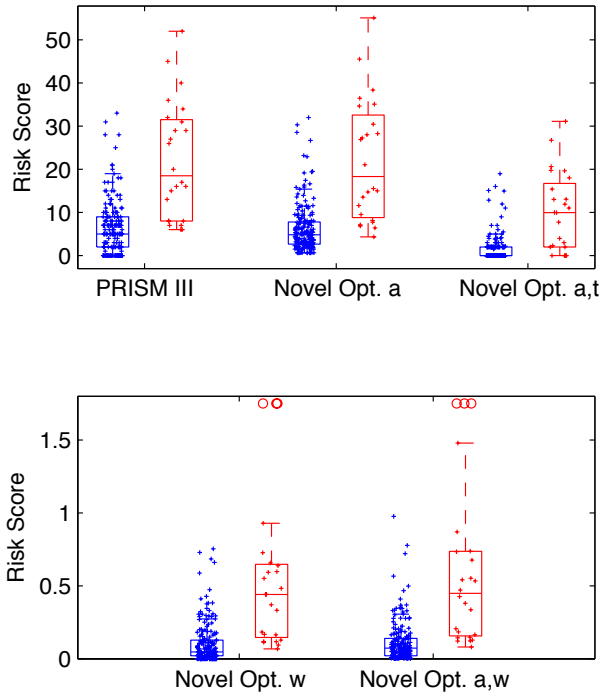

Fig. 2 Risk scores for survivors (blue) and nonsurvivors (red). The center mark on the box indicates the median while the edges of the box mark the 25th and 75th percentile. Individual scores are plotted as +. In the bottom panel, the red circles indicate subjects with scores higher than 2.

zero contribution to the risk score for all patients. This optimization results in higher accuracy than PRISM III in the upper range of scores (higher J and precision-recall balance) and poorer classification than PRISM in the lower score range (lower AUC). However, we expect that these results would improve with a larger patient population given the age-dependency of some of thresholds and the mortality distribution across age groups (Table 2). The ROC results for PIM2 are also included in Table 3 to compare the novel score to another widely used pediatric risk score. The novel score optimized over any of the parameters outperforms PIM2 in terms of AUC, and optimization over 2 parameters also yields a higher precision-recall balance than PIM2. Though PIM2 results in a slightly higher J than the novel score optimized over $\boldsymbol{a}, \mathbf{w}$, the large gain in AUC and higher precision-recall balance make the novel score the preferred choice for predicting risk of in-hospital mortality in the studied population.

The boxplots in Figure 2 illustrate how the proposed prediction scheme compares with PRISM III. Despite additional features used by our scheme (due to splitting of the OR statements into components), the novel scores have similar average values to those of PRISM III. Given the class imbalance,

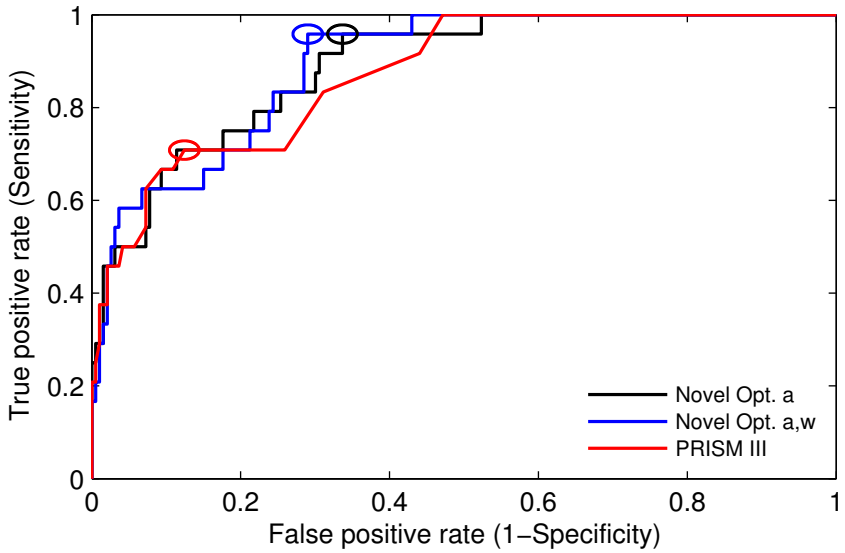

Fig. 3 ROC curve for PRISM III (red) and novel scores optimized over the nonlinear transformation slopes $\boldsymbol{a}$ (black) and alternating minimization over the slopes $\boldsymbol{a}$ and weights $\mathbf{w}$ of the nonlinear transformations (blue). The circled points correspond to those that maximize Youden's index for each curve.

this is likely the result of lower slopes in the nonlinear transformations which decrease the feature contributions of survivors having measurements near the soft thresholds. The movement of soft thresholds outside the physiological range following inclusion of $\mathbf{t}$ as an optimization parameter results in lower average score values compared to PRISM III. Finally, reduction in mean scores when the weights are included as optimization variables is expected due to the prior on the weight distribution.

The difference in AUC between PRISM III and the novel scores (Figure 3) is primarily caused by low PRISM III scores for some nonsurvivors. Due to the finite set of PRISM III values, patients are more likely to share the same score which causes decrease in both the true positive rate and the false positive rate as the cutoff value is lowered in the ROC analysis. The largest differences in the curves can be traced to patients with PRISM III scores between 7 and 13. While 45 out of 193 survivors have PRISM III scores in this range, so do 7 out of 24 nonsurvivors.

To test the advantage of the nonlinear transformations of the features, we performed a logistic regression with ridge, lasso (Tibshirani, 1994), and elastic net (Zou and Hastie, 2005) penalties on the age and raw variables of the datasets. Since missing data is prevalent and complete data sets are required for logistic regression, missing values of the variables are imputed by k nearest-neighbors (kNN) (Hastie et al, 1999), the probabilistic principal components analysis (PPCA) method (Tipping and Bishop, 2002), mean values and normal

Table 4 ROC: Logistic with raw variables vs. PRISM with imputed data

		kNN	PPCA	mean	normal
AUC	PRISM III	0.8790	0.8741	0.8709	0.8735
	Raw	0.8437	0.8683	0.8400	0.8141
J	PRISM III	0.6047	0.6153	0.5788	0.5840
	Raw	0.5479	0.5889	0.5637	0.5682
PrecRec	PRISM III	0.5000	0.5000	0.5172	0.5172
	Raw	0.5000	0.5542	0.5833	0.5833

values (Behrman et al, 2004). In kNN imputation, Euclidean distances normalized by the number of common features between patients are calculated and missing values are imputed as the average value of the k-nearest neighbors that observe the variable. PPCA aims to reduce the dimensionality of the data by associating a Gaussian latent variable model with the observed data and imputing missing values by an iterative expectation-maximization procedure. Data were imputed with $k = 5$ and 4 principal components. The best values for AUC were achieved with the elastic net and are compared to the PRISM III values calculated with the imputed dataset in Table 4.

Though logistic regression with variables imputed with PPCA results in a slight improvement over the PRISM III scores without imputation in terms of J and the precision-recall balance, the PRISM III scores calculated with the imputed data outperforms both logistic scores in terms of AUC and J. Logistic regression with variables imputed with mean or normal values results in the same PrecRec value as the novel score. However, the novel score greatly outperforms both of the raw variable models in terms of AUC and J. This suggests the variables in the scores should be nonlinearly transformed for optimal mortality prediction. Note that the novel score without imputation performs better than the scores presented in Table 4 in terms of all of the evaluation criteria.

5 Conclusion

We have developed a novel outcome prediction score that exploits advantages of PRISM III and addresses a key limitation, resulting in a significantly more accurate predictor of risk of in-hospital mortality in children admitted to PICU. In particular, by transforming predictive variables using a combination of logistic functions, the developed method allows for a fine differentiation between critical and normal values of the predictive variables. Optimization of the continuous score allows for not only specifying different weights for the variables but, by optimizing over the slope and/or inflection point of the logistic curve in the feature transformation, we can also identify the range of values of each variable where the risk increases. This optimization need only be performed once to determine the optimal parameters and the score is thereafter

quickly calculated for each patient. Optimal values of the parameters of logistic functions may be readily re-learned as the patient population and standards of care evolve. In a study of brain trauma pediatric patients, the new score demonstrated significantly higher predictive power than PRISM III.

The presented method that can be broadly applied to devise and optimize risk scores with predictive power superior to schemes that use hard-thresholding of physiological variables. A potential future application of the developed method is optimization of in-hospital mortality risk scores designed for adult ICU population, such as the Acute Physiology and Chronic Health Evaluation (APACHE) (Knaus et al, 1991) and the Simplified Acute Physiology Score (SAPS) (Moreno et al, 2005). Both of these scores are calculated by adding points whenever physiologic variables are in a given range. Using the proposed method, not only can scores be calibrated by assigning appropriate variable weights, but also the variable ranges where the risk increases can be learned for specific patient populations.

Acknowledgment

The authors would like to thank Karen Piper for the identification of patients that met the inclusion criteria for the study.

References

Bahloul M, Chaari A, Chabchoub I, Medhyoub F, Dammak H, Kallel H, Ksibi H, Haddar S, Rekik N, Chelly H, Bouaziz M (2011) Outcome analysis and outcome predictors of traumatic head injury in childhood: Analysis of 454 observations. Journal of emergencies, trauma, and shock 4(2):198–206

Baker SP, O'Neill B, Haddon J W, Long WB (1974) The injury severity score: a method for describing patients with multiple injuries and evaluating emergency care. The Journal of Trauma 14(3):187–196

Behrman RE, Kiegman RM, Jenson HB (eds) (2004) Nelson Textbook of Pediatrics, edition 17 edn. Elsevier

Bhadoria P, Bhagwat AG (2008) Severity scoring systems in paediatric intensive care units. Indian Journal of Anaesthesia 52(Suppl (5)):663–675

Boyd S, Vandenberghe L (2009) Convex optimization. Cambridge Univ Press

Brady AR, Harrison D, Black S, Jones S, Rowan K, Pearson G, Ratcliffe J, Parry GJ (2006) Assessment and optimization of mortality prediction tools for admissions to pediatric intensive care in the united kingdom. Pediatrics 117(4):e733–742

Burd RS, Madigan D (2009) The impact of injury coding schemes on predicting hospital mortality after pediatric injury. Academic Emergency Medicine 16(7):639–645

Cantais E, Paut O, Giorgi R, Viard L, Camboulives J (2001) Evaluating the prognosis of multiple, severely traumatized children in the intensive care unit. Intensive Care Medicine 27(9):1511–1517

Courville XF, Koval KJ, Carney BT, Spratt KF (2009) Early prediction of posttraumatic in-hospital mortality in pediatric patients. Journal of Pediatric Orthopedics 29(5):439–444

Gemke RJ, van Vught J (2002) Scoring systems in pediatric intensive care: PRISM III versus PIM. Intensive care medicine 28(2):204–207

Goddard JM (1992) Pediatric risk of mortality scoring overestimates severity of illness in infants. Critical care medicine 20(12):1662–1665

Hastie T, Tibshirani R, Sherlock G, Eisen M, Brown P, Botstein D (1999) Imputing missing data for gene expression arrays. Tech. rep., Stanford University

Hilden J, Glasziou P (1996) Regret graphs, diagnostic uncertainty and youden's index. Statistics in medicine 15(10):969–986

Karambelkar G, Mane S, Agarkhedkar S, Karambelkar R, Singhania S, Kadam S (2012) The relevance of 24 hour prism iii score in predicting mortality in pediatric intensive care unit. Int J Pharm 3(4):214–219

Knaus W, Wagner D, Draper E, Zimmerman J, Bergner M, Bastos P, Sirio C, Murphy D, Lotring T, Damiano A (1991) The APACHE III prognostic system. risk prediction of hospital mortality for critically ill hospitalized adults. Chest 100(6):1619–1636

Laurer HL, Meaney DF, Margulies SS, Mcintosh TK (2002) Modeling brain Injury/Trauma. In: Ramachandran V (ed) Encyclopedia of the Human Brain, Academic Press, New York, pp 93–102

Maas AIR, Harrison-Felix CL, Menon D, Adelson PD, Balkin T, Bullock R, Engel DC, Gordon W, Langlois-Orman J, Lew HL, Robertson C, Temkin N, Valadka A, Verfaellie M, Wainwright M, Wright DW, Schwab K (2011) Standardizing data collection in traumatic brain injury. Journal of neurotrauma 28(2):177–187

Marcin JP, Pollack MM (2000) Review of the methodologies and applications of scoring systems in neonatal and pediatric intensive care. Pediatric Critical Care Medicine 1(1):20–27

Marcin JP, Pollack MM (2002) Triage scoring systems, severity of illness measures, and mortality prediction models in pediatric trauma. Critical Care Medicine 30(11 Suppl):S457–467

Martha VF, Garcia PCR, Piva JP, Einloft PR, Bruno F, Rampon V (2005) Comparison of two prognostic scores (PRISM and PIM) at a pediatric intensive care unit. Jornal de pediatria 81(3):259–264

Moreno R, Metnitz P, Almeida E, Jordan B, Bauer P, Campos R, Iapichino G, Edbrooke D, Capuzzo M, Le Gall JR (2005) SAPS 3—From evaluation of the patient to evaluation of the intensive care unit. part 2: Development of a prognostic model for hospital mortality at ICU admission. Intensive Care Medicine 31(10):1345–1355

Perkins NJ, Schisterman EF (2006) The inconsistency of "optimal" cutpoints obtained using two criteria based on the receiver operating characteristic curve. American journal of epidemiology 163(7):670–675

Pivovarov R, Elhadad N (2012) A hybrid knowledge-based and data-driven approach to identifying semantically similar concepts. Journal of Biomedical

Informatics 45(3):471–481
Pollack MM, Patel KM, Ruttimann UE (1996) PRISM III: an updated pediatric risk of mortality score. Critical Care Medicine 24(5):743–752
Ponce-Ponce De León AL, Romero-Gutiérrez G, Valenzuela CA, González-Bravo FE (2005) Simplified PRISM III score and outcome in the pediatric intensive care unit. Pediatrics international 47(1):80–83
Qureshi AU, Ali AS, Ahmad TM (2007) Comparison of three prognostic scores (PRISM, PELOD and PIM 2) at pediatric intensive care unit under pakistani circumstances. Journal of Ayub Medical College, Abbottabad: JAMC 19(2):49–53
Royston P, Altman DG, Sauerbrei W (2006) Dichotomizing continuous predictors in multiple regression: a bad idea. Statistics in Medicine 25(1):127–141
Ruttimann UE, Pollack MM (1993) A time-series approach to predict outcome from pediatric intensive care. Computers and Biomedical Research, an International Journal 26(4):353–372
Saria S, Rajani AK, Gould J, Koller D, Penn AA (2010) Integration of early physiological responses predicts later illness severity in preterm infants. Science Translational Medicine 2(48):48ra65–48ra65
Scavarda D, Gabaudan C, Ughetto F, Lamy F, Imada V, Lena G, Paut O (2010) Initial predictive factors of outcome in severe non-accidental head trauma in children. Child's Nervous System: ChNS 26(11):1555–1561
Shann F, Pearson G, Slater A, Wilkinson K (1997) Paediatric index of mortality (PIM): a mortality prediction model for children in intensive care. Intensive Care Medicine 23(2):201–207
Slater A, Shann F, Pearson G (2003) PIM2: a revised version of the paediatric index of mortality. Intensive Care Medicine 29(2):278–285
Slater A, Shann F, ANZICS Paediatric Study Group (2004) The suitability of the pediatric index of mortality (PIM), PIM2, the pediatric risk of mortality (PRISM), and PRISM III for monitoring the quality of pediatric intensive care in australia and new zealand. Pediatric Critical Care Medicine 5(5):447–454
Streiner DL (2002) Breaking up is hard to do: The heartbreak of dichotomizing continuous data. Canadian Journal of Psychiatry 47(3):262
Sun J, Hu J (2012) Combining knowledge and data driven insights for identifying risk factors using electronic health records. AMIA Annual Symposium proceedings / AMIA Symposium AMIA Symposium 2012:901–10
Thukral A, Lodha R, Irshad M, Arora NK (2006) Performance of pediatric risk of mortality (PRISM), pediatric index of mortality (PIM), and PIM2 in a pediatric intensive care unit in a developing country. Pediatric Critical Care Medicine 7(4):356–361
Tibby S, Taylor D, Festa M, Hanna S, Hatherill M, Jones G, Habibi P, Durward A, Murdoch I (2002) A comparison of three scoring systems for mortality risk among retrieved intensive care patients. Archives of Disease in Childhood 87(5):421–425
Tibshirani R (1994) Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society, Series B 58:267–288

Tipping ME, Bishop CM (2002) Probabilistic principal component analysis. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 61(3):611–622

Volakli EA, Sdougka M, Drossou-Agakidou V, Emporiadou M, Reizoglou M, Giala M (2012) Short-term and long-term mortality following pediatric intensive care. Pediatrics international: official journal of the Japan Pediatric Society 54(2):248–255

Wells M, Riera-Fanego JF, Luyt DK, Dance M, Lipman J (1996) Poor discriminatory performance of the pediatric risk of mortality (PRISM) score in a south african intensive care unit. Critical care medicine 24(9):1507–1513

Zou H, Hastie T (2005) Regularization and variable selection via the elastic net. Journal of the Royal Statistical Society, Series B 67:301–320