Prognosis-based Decision Support in Medicine A Divide and Conquer Approach

Ameen Abu-Hanna

Department of Medical Informatics, AMC-University of Amsterdam
Meibergdreef 15, 1105 AZ Amsterdam, The Netherlands

Abstract. Managers of clinical departments are constantly seeking ways to assess and improve their quality of care. In doing so they feel the need for decision support methods and tools. "Prognosis-based decision support" is an elemental way to achieve support where a model is learned for the prediction of a patient outcome indicative of care quality. The idea is that quality of care can be assessed by comparing the model's predictions for a patient population to the actual outcomes of these patients, and hence facilitate making decisions about future improvement of care delivery. This paper first describes the current main stream prognosis-based decision support process which is centered around a global logistic model. It then suggests an improved, more transparent method based on learning prognostic models that identify patient sub-populations (using a classification tree) for which specialized prognostic models (based on local regression) are built. The method is illustrated in the field of intensive care.

1 Introduction

Health Care is expensive and evidence for its effectiveness and efficiency is continuously sought. Many national and international quality programs have been set up to assess health care quality. For example our department is responsible for national registries such as those for intensive care and cardiology and international ones such as that for the European Renal Association. These registries include information about patients and outcomes of care such as mortality and co-morbidity. When the outcome is indicative of the quality of care, registries form valuable vehicles for analytic instruments aimed at the evaluation of the quality of care and supporting decisions for its constant improvement.

Prognostic models for outcome prediction form indispensable ingredients among these evaluative tools [1, 15] enabling what can be termed as "prognosis-based decision support". Intensive care (IC), which will be used throughout the paper for illustrative purposes, is a good example of this. Here various prognostic models, such as the SAPS-II [13], version II of the Simplified Acute Physiology Score, have been developed to estimate the probability of in-hospital mortality. In-hospital mortality concerns deaths in the hospital during or after stay in an Intensive Care Unit (ICU). Like many other prognostic models in medicine these are statistical models that can be characterized by their use of a small

set of covariates, at the heart of which is a *score* variable reflecting the patient's overall severity of illness, and by their reliance on the probabilistic logistic model. The common practice in decision support is to arrange predicted probabilities of death in groups, e.g. quantiles, and compare each group's mean with the fraction of observed deaths corresponding to that group. Discrepancies are then assessed by experts to find out whether actions to improve care delivery should be taken.

This paper aims at enhancing the decision support process by introducing the notion of patient sub-population in the process: Instead of using one global logistic regression model, a classification tree is induced to devise patient groups for which local models, both parametric and nonparametric, can be fit. The idea is twofold: first, clinicians are used to think in terms of sub-groups and hence stratification on these groups brings transparency allowing for focusing on them and, second, the method allows for an ensemble of models that together can fit the data better than one global parametric model.

A balance must be sought between the ability to divide patients into possibly many sub-groups to enhance focus on their physiological homogeneity on the one hand, and on the other hand, lumping up patient descriptions based on the more abstract notion of overall severity of illness that avoids fragmentation and allows for parametric models that assume monotonicity (higher score implies higher probability of death). Our approach to this problem is to devise patient sub-populations, using a classification tree, based on the physiological variables that *underlie* the severity of illness score while using the overall lumped score itself in the local prognostic models. The hypothesis is hence that the severity of illness score holds additive value that can be tapped and made explicit by the tree. The number of sub-groups can be controlled by pruning the tree at places where the local models fail to enhance prognostic performance. We show in this paper that in measuring prognostic performance for decision support in a quality of care program, the popular error rate and even area under the Receiver Operator Characteristic (ROC) curve–which are measures of *accuracy*– play a relatively minor role in the evaluation of models and we explain why measures of *precision* are usually more appropriate.

The paper is organized as follows. Prognosis-based decision support is introduced in Sect. 2 and illustrated in intensive care. Then in Sect. 3 aspects of model evaluation are provided where it is explained why measures of precision are important in decision support of quality of care programs. Sect. 4 presents our method for improved decision support based on a hybrid model and is illustrated in two variants. Sect. 5 concludes this paper.

2 Prognosis-based DS for Quality of Care

The need for assessing and assuring the quality of care in medicine arises from various reasons. In the intensive care the technologies to treat and monitor organ failure of critically ill patients are not only costly as consequence of the price tags of the advanced equipment, they also require a large number of skilled personnel to maintain and use these technical facilities 24 hours a day, 7 days a week.

In addition, differences in the view on the best delivery of care, professional ambitions, budgetary constraints and insurance regulations now have prompted physicians and managers to assess the quality of IC treatment.

One way to achieve objective appraisal of the quality of the care process, especially if experimentation is impossible or unethical, is prognosis-based: outcome data such as mortality, morbidity and length of stay are compared to *predicted* outcome values. This is practiced in various programs such as those described in [18, 10]. The predictions should take into account the characteristics of the patient population admitted to an ICU. This is called case-mix adjustment and is usually quantified by a score of the severity of illness of each patient at the time of admission. In the main, existing IC case mix adjustment models are concerned with predicting mortality. Prognostic models that make these case-mix adjusted predictions lie at the heart of quality assessment efforts.

2.1 Current Practice using Logistic Regression

Two major prognostic systems in IC are the Acute Physiology and Chronic Health System, APACHE [11] and the Simplified Acute Physiology Score, SAPS [13]. Both come in different versions (e.g. I and II). These systems, like many others, are based on logistic regression.

A logistic regression model is a parametric model that specifies the probability of a dichotomous variable Y having the value 1 –indicating the occurrence of an event such as death– given the values of the covariates of the model. It has the following form:

$$p(Y = 1|\boldsymbol{x}) = \frac{e^{g(\boldsymbol{x})}}{1 + e^{g(\boldsymbol{x})}} \tag{1}$$

where $\boldsymbol{x} = (x_1, ..., x_m)$ is the covariate vector. For m variables (also called predictors) the *logit function* $g(\boldsymbol{x})$ has the following form:

$$g(\boldsymbol{x}) = \beta_0 + \sum_{i=1}^{m} \beta_i x_i \tag{2}$$

where β_i, $i = 1, ..., m$, denote the coefficients of the m predictors. Fitting a logistic regression model implies finding estimates of the $\beta_0, ..., \beta_m$ that maximize the likelihood of the model (that is, the probability of the data given the model). Note that the conditional probability in Eq. (1) is in effect the conditional expectation $E(Y = 1|\boldsymbol{x})$ because Y is binary.

The monotonicity of logistic regression means that it is not a good idea to represent a raw physiological feature such as temperature of the patient directly as a covariate because an increase e.g. from 35 to 37 Celsius degrees should have an opposite effect on the probability of the adverse event than an increase from, say, 37 to 39 Celsius degrees. Because of this and due to other reasons,

the APACHE and SAPS use one or more aggregate *scores* for the quantification of the severity of illness as covariates instead of the individual features. A higher score corresponds to greater deviation from the healthy status and hence a higher probability of death. Different elements contribute to this total additive score such as physiological variables e.g. heart rate, and white blood cell count; demographic variables e.g. age; and covariates concerning earlier chronic history. For example the total SAPS-II score ranges from 0 to 163 points and is related to a probability of hospital mortality based on the following logit model:

$$g(score) = -7.7631 + 0.0737\, score + 0.9971\, ln(score + 1)$$

Such models are used by quality of care programs such as the National Intensive Care Evaluation (NICE) program. NICE is a foundation established in 1996 by an initiative of a professional group of intensivists to gain insight and to improve the effectiveness and efficiency of Dutch intensive care units (ICUs). Many different ICUs in the Netherlands are currently participating in NICE. The following is a sketch of the procedure followed by NICE in supporting decisions about quality of care.

1. Each participating ICU provides its care-related data following a standard format respecting agreements which are semi-formalized in a data dictionary. For each patient more than 200 items are collected including patient description and various outcomes including length of stay and mortality.
2. The information from all participants is accumulated and stored in the NICE registry after data quality validation procedures have taken place.
3. Based on data from all ICUs various prognostic models are constructed including SAPS and APACHE.
4. Each prognostic model is validated on a large test set that includes data from all participating ICUs. The Hosmer-Lemeshow tests for goodness of fit are usually used in this step which imply the comparison between predicted probabilities and observed proportions of mortality within various probability groups.
5. Once a prognostic model is found satisfactory it may be considered as predicting the "average outcome of a national" ICU. It is then used to predict probability outcomes of new data from each ICU. For *each* ICU the model's predictions are once again lumped into probability groups and compared with the actual proportion of mortality in that ICU.
6. Representatives of the participating units meet periodically and discuss discrepancies between their "performance" and the "national average" predictions for their patients. The reasons for the discrepancy are sought: is there a discrepancy in definitions of some data items? is the model to be trusted for a specific risk group? is there something that can be learned from those ICUs that are faring better than average?
7. Based on the outcomes of the last step, decisions are made for future improvement of care.

3 Model Evaluation Aspects

Before we present our method for constructing prognostic models for improved decision support a word is due about model evaluation aspects in quality of care programs. Consider two different models M1 and M2 that when given a specific score, $score_0$, as the covariate value they provide two different mortality probability estimates, say .55 and .85. If a prognosis is sought for *individual* patients with $score_0$, both models will predict non-survival for each of these patients (because the probability assigned for the event –non-survival– is greater than the probability assigned for survival). Assuming there are more non-survivors than survivors among the patients with $score_0$, these two models are equally effective in discriminating between the two survival states and hence will inflict the same total error rate (the Bayes error rate) in this group.

There is however a substantial difference between the probability estimates of the two models which cannot be ignored if one is interested in the *probability* of an event within a *group* of patients. To judge the performance of an ICU, one compares the *estimated probability* of non-survival with the *observed mortality rate* of a specific group. If 70% of patients with $score_0$ did not survive then according to M1 the ICU is performing very poorly but according to M2 it is performing much better than expected for that group of patients. We are hence interested in a *precise* model that provides honest estimates of the *true probability* of an event rather than merely a discriminating model with the ability to assign the highest probability to the actual event or class. However, error rate still forms one of the most used measures for evaluating classifiers (in machine learning and statistics) [7]. Along with other measures of accuracy, such as ROC (Receiver Operator Characteristic) and the aggregated area under its curve (AUC), these measures are useful but alone do not tell the whole story. See [7] for a comprehensive framework of evaluation methods and measures.

Current main stream evaluation of logistic regression models such as the APACHE-II and SAPS-II models usually rely on the Hosmer-Lemeshow statistics [8] where it is referred to as calibration (see [17] for a more general discussion). These are essentially precision measures that group predicted probabilities in *non-overlapping* regions. A major disadvantage of the Hosmer-Lemeshow statistics is that they have been shown to be quite sensitive to the cut-off points used to form the expected probability regions [9].

In this work we used various measures of accuracy and precision in order to inspect the performance of prognostic models. These include the Brier score, also known as the mean quadratic error as a measure of accuracy. For precision we obtain information about the "true" probability from the *test* set in two ways: A direct one where local regression is used to smooth mortality data where subsequently the quadratic error of the predicted ones is used as the (im)precision measure. In a second indirect way we compare a transformation *summary* obtained from the predicted probabilities on the test set to a similar summary obtained from the true observed outcomes from that test set. The difference between the summaries leads to a statistic that is significantly different

from 0 when precision is bad. We use the logarithmic transformation [7] to obtain each summary.

Our precision measures do not require division in non-overlapping probability groups in the expected probability of mortality, and hence do not share the drawbacks of the Hosmer-Lemeshow statistics mentioned in [9]. Our precision measures rely on some smoothing aspect of the neighborhood of a point of interest. This neighborhood is dynamic and overlaps with other neighborhoods. Moreover our measures will be obtained on patient groups sharing various inherent characteristics (e.g. physiological variables) as will be demonstrated below.

4 Suggested Improved Method: divide and conquer

In this section we motivate the use of a method aimed at a better understanding of the IC data and the models fitted to it in order to enhance decision making. In this method we try to exploit information that is implicit in the severity of illness score covariate(s). For illustration purposes we will concentrate on a model that uses only one score as in the SAPS model. As an example of the information that may be masked by a score consider that a patient with unscheduled surgery (8 points) and normal heart rate (0 points) will score as a patient with a medical admission (6 points) with a heart rate between 40 and 69 (2 points), assuming for simplicity that they score the same for the other attributes. Our approach is to create sub-groups that share some of their underlying attribute values.

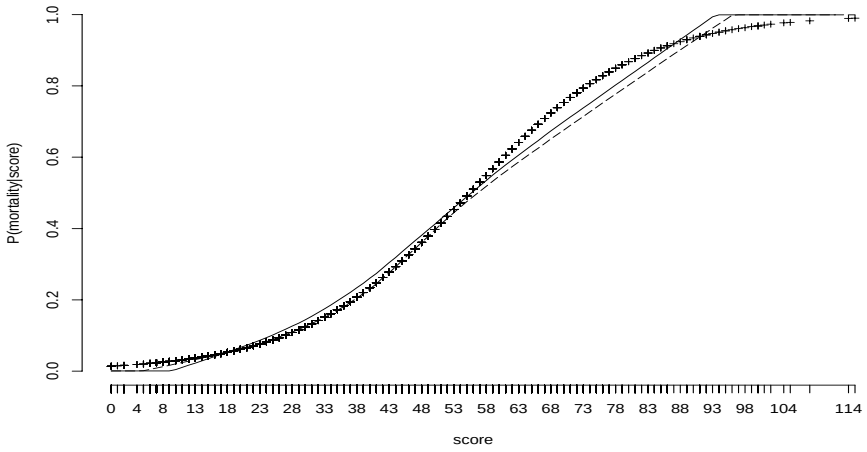

Fig. 1. Predictions of mortality on a test set by logistic regression (points marked with +) and Lowess (solid line). The dashed line represents Lowess smoothing on the test set itself.

A classification tree is used[1] to create these groups while using mortality as the classification attribute. The idea is that creating groups (the divide step) with different distribution of the mortality attribute and fitting local models (the conquer step) on them will prove advantageous if compared to the original global model. The total score however will be the sole covariate in the local models fitted once a group is identified. For the local models we have experimented with both (parametric) logistic regression models and nonparametric local regression models based on Cleveland's Lowess procedure [6]. As an example consider the predictions of two models (trained previously on a separate training set of some similar patient group) in Fig. 1: the predictions are those of Logistic regression (with its characteristic S-shape), and the non-parametric local regression (the solid line) according to Lowess. The dashed line represents Lowess smoothing on the test set itself in order to provide an idea of the "true" mortality function. Note how the two models over or underestimate this "true" probability in this test set.

Table 1. Characteristics of important attributes in the dataset.

Variable name	*Description*		*Mean±s.d.*	*Normal range*
syst.min	minimal systolic blood pressure		92.0±32.4	100-199
urine.24	urine production in first 24 hrs		2.6±2.3	>1
heartr.min	minimum heart rate		71.2±23.0	70-119
bicarb.min	minimum bicarbonate		22.4±5.2	≥20
bicarb.max	maximum bicarbonate		25±4.5	≥20
gcs.low	lowest Glasgow Coma Scale		13.8±3.2	15
wbc.max	maximum white blood cell count		12.1±11.4	1-19.9
Variable name	*Description*	*value*	*Freq.*	
adm.type	admission type			
	medical admission	1	45.3%	
	unscheduled surgical	2	17.7%	
	scheduled surgical	3	37.0%	

As an illustration of the method, consider the simple global logistic model, termed, N-SAPS (the N of NICE), that was developed on a training set of 5218 IC admissions of NICE, using only the SAPS-II score as covariate (without the logarithm of the score, in order to concentrate only on the score). This model is later evaluated on a separate test set of 2585 admissions. The N-SAPS model has the following form:

$$g(score) = -4.2548 + 0.0737\, score$$

[1] We use the *rpart* function available in the Splus statistical package which is an implementation of CART [4].

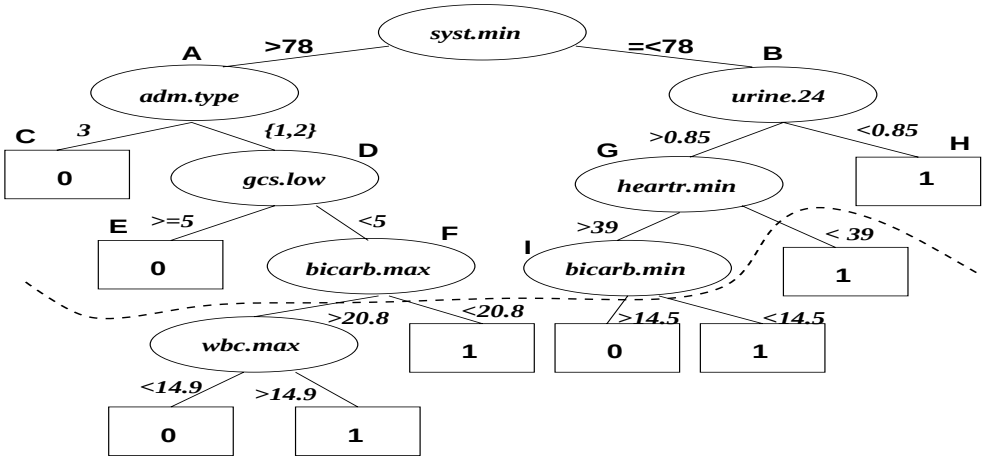

Fig. 2. The IC classification tree based on physiological attributes and admission type, see Table 1 for attribute descriptions. The leaves denote the majority class. The nodes above the dashed line constitute the identified groups with sufficient size.

N-SAPS was then compared with a model constructed according to our method whose classification tree appears in Fig. 2. Attribute names are explained in Table 1. The restriction to a binary tree is aimed at combating fragmentation and because all attributes are either continuous or ordinal (admission type can be clinically viewed this way in the sense that higher values of admission type indicate more serious conditions). Information gain based on entropy was used as the criterion for the selection of attributes in the tree. Note that patients with the same score could end up in different nodes in this tree as most scores can be obtained by different combinations of e.g. physiological variables.

When viewing the probability functions in the different nodes on the training set by smoothing the raw mortality data (using Lowess) there were obvious differences between them. The functional form at each node turned out to be *qualitatively consistent* in random samples of the data set, as long as there were sufficient instances in each node. This suggests that the tree is exploiting information which has not been explicitly used in the score and hence is masked from N-SAPS. One way to use this insight in decision support is simply to induce the tree partition in data from a different ICU, or data from the same ICU but taken at a different time and inspect the conformance to this qualitative behavior to detect differences.

When formally inspecting the model performance by obtaining our performance measures for the prognostic models on the tree nodes with at least 100 cases (the nodes appearing above the dashed line in Fig. 2) the following results were obtained (see Tables 2 and 3): The N-SAPS model has been outperformed in accuracy and precision in most of the cases meaning that the attributes underlying the score variable do have an added value. The local Lowess models did not outperform the local parametric logistic regression model indicating perhaps

Table 2. Accuracy measures for the three models (values have been multiplied by $*10^3$). Bold values mean accuracy is significantly better than in the N-SAPS model.

TLR and TLW stand for the local logistic and Lowess models.

		Node and ♯instances in test set								
		A	B	C	D	E	F	G	H	I
Model	*Msr*	2072	513	900	1172	1063	109	402	111	372
N-SAPS	*Brier*	104.67	173.96	50.106	146.56	137.01	239.74	183.57	139.17	176.04
TLR	*Brier*	104.62	174.14	**49.73**	146.12	136.82	238.91	182.36	**126.17**	174.83
TLW	*Brier*	105.01	175.15	**49.861**	146.53	137.40	**231.74**	183.17	128.08	175.44

Table 3. Direct and indirect precision measures (values have been multiplied by $*10^3$). Values in bold mean they are significantly better than the N-SAPS model values. Boxed values mean statistically significant bad precision.

		Node and ♯instances in test set								
		A	B	C	D	E	F	G	H	I
Model	*Msr*	2072	513	900	1172	1063	109	402	111	372
N-SAPS	*direct*	0.59	0.52	1	1.1	0.52	16.42	2.1	10.1	1.75
	indirect	-987.57	613.81	-3107.4	1725.1	1175.5	2528.5	1127.6	-785.28	264.57
TLR	*direct*	**0.57**	**0.34**	**0.85**	**0.35**	0.47	**15.95**	**0.4**	**0.42**	**0.35**
	indirect	-694.4	-42.701	-537.47	-265.71	-635.82	2051	-243.46	-76.171	-784.05
TLW	*direct*	**0.42**	**0.2**	**0.88**	**0.4**	**0.3**	**11**	**0.56**	**0.91**	**0.73**
	indirect	-1397.7	478.68	-1147.4	-407.25	-574.09	1633.4	398.05	-442.94	-290.18

that the nodes do not contain sufficient instances to allow for a complete non-parametric model. The non-parametric model is however quite useful in order to inspect the qualitative shape of the mortality function.

When inspecting the results were N-SAPS could be improved, the following sub-populations are notable:

Node **C** Patients with a normal to high systolic blood pressure who are admitted after scheduled surgery; these patients are relatively healthy.

Node **F** Patients with a normal to high systolic blood pressure who are admitted after unscheduled surgery or for a non-surgical (medical) reason and who have a very disturbed neurological status reflecting a high severity of illness.

Node **H** Patients with low blood pressure and low urine production reflecting possible heart and renal failures. These patients are seriously ill.

This result is in line with other studies in which the precision of current (global) logistic regression models is often found lacking at the extremes: patients which are relatively healthy and patients which are seriously ill. The difference is however that in our analysis one can characterize these groups in terms of their attribute values instead of establishing that the model does not calibrate for very low and very high scores where the different patient populations cannot be further discerned.

4.1 Deviation-based Trees

From an epidemiological point of view it is interesting for decision makers to communicate about patients using the notion of risk factors. We can employ our method to identify important risk factors that complement the severity of illness score that can be viewed as an overall risk factor. Instead of building trees based on the underlying physiological and other raw attributes one can use attributes that pertain to *deviation* from the values in the normal ranges. For example instead of using the ranges of heart-rate we use the "penalty" scores for heart-rate as a co-variate in the induction of the classification tree.

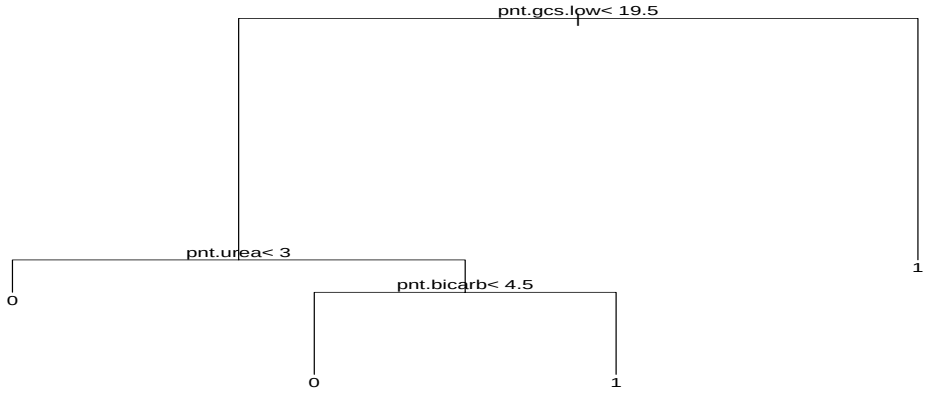

Fig. 3. Classification tree based on deviation, the prefix "pnt" indicates (deviation) points of the corresponding attribute. Each condition corresponds to the *left* branch. The 0 and 1 represent survival and death respectively.

To illustrate, consider Fig. 3 that shows a deviation-based tree induced from a similar intensive care training set (under slightly different circumstances). As pointed out in [16] one should be careful in interpreting risk factors. In fact due to the meaning of the new attributes as adversary conditions, any path that mixes a greater-than (">") condition (meaning worse condition) together with a less-than ("<") condition should be interpreted with care. For example, in the high risk group "$pnt.tgc.low < 19.5$ & $pnt.urea \geq 3$ & $pnt.bicarb \geq 4.5$" one should not conclude that the low value of *pnt.gcs.low* itself is associated with high risk. The fact that low *pnt.gcs.low* appears in the path of a high risk group is not by definition an indication of an interaction with the other factors in the path but could be simply indicative of being used as a "rest" category that the tree has created when using high *pnt.gcs.low* for constructing the first group.

This is a phenomenon introduced by the way trees are induced of which decision makers must be aware.

5 Conclusions

From this work, one can postulate that using the hybrid method of a classification tree with local prognostic parametric and non-parametric models provide better insight into the data and hence enhances decision support. One may conclude that the attributes underlying severity of illness scores can indeed contribute to better models. It is interesting to note that in our search for a balance for granularity of covariates we started from the severity of score which turns out to be too lumped, while in other approaches such as [3] one seeks a balance by searching for aggregations of (too) low level features.

The idea of comparing logistic regression with classification trees (e.g. as done in [14]) or the combination of a classification tree with other models are in themselves not new. In [12], for example, Naive Bayes Classifiers are fit on some of the tree nodes to boost classification. Recently a related idea to ours has been proposed in [5] for learning "Treed Models" based on Bayesian methods where also logistic regression models have been proposed. The contribution of our work lies in providing a motivated synthesis of modeling and evaluation concepts tailored to the specific constraints of decision support in quality of care programs with emphasis on precision. This emphasis on precision would, for example, not allow for models such as Naive Bayes Classifiers which are known, in the main, to have good and robust classification error but are often quite imprecise.

Further work includes putting our method to the test in the decision support process to see how it affects the intensivists' decision making. Another topic is to empirically investigate when to stop growing the tree and how to opportunistically combine all (here three) model types for providing the best prognosis. An important sequel to the work presented in this paper is the treatment of other outcome measures than mortality, such as length of stay in the ICU and organ failure scores registered daily for each patient, that although more complex to handle they do provide more sensitive measures of quality of care.

One might conclude that our results provide proof of concept that in this era of large amounts of electronic data, there is room for a variety of modeling concepts for enhancing decision support. We feel that the inspection of interesting patient sub-populations is an important enrichment to the traditional logistic regression models.

Acknowledgment Thanks are due to the board of the National Intensive Care (NICE) foundation for its support and feedback. The board consists of: G.J. Scheffer, R. Bosman, E. de Jonge, J.C.A. Joore, N.F. de Keizer, H.H.M. Korsten, J.G. van der Hoeven, P.H.J. van der Voort. Furthermore, we are grateful to all NICE participants for collecting the data. Special thanks to Nicolette de Keizer for feedback on this work.

References

1. Abu-Hanna A and Lucas PJF. Prognostic Models in Medicine AI and Statistical Approaches, (Abu-Hanna A. and Lucas PJF, eds.). Special issue of Methods of Information in Medicine 2001, 40:1–5.
2. Abu-Hanna A and de Keizer N. A Classification-Tree Hybrid Method for Studying Prognostic Models in Intensive Care, AIME 2001, 99–108.
3. Bohanec M, Zupan B, Rajkovic V. Applications of Qualitative Multi-attribute Decision Models in Health Care. International journal of Medical Informatics. 2000, vol. 58-59, 191–205.
4. Breiman L, Friedman JH, Olshen RA, Stone CJ. *Classification and Regression Trees.* Belmont: Wadsworth, 1984.
5. Chipman H, George E, McCulloch R. Bayesian Treed Models. Machine Learning, 48, 299-320, 2002.
6. Cleveland WS. Robust Locally Weighted Regression and Smoothing Scatterplots. J. Amer. Statist. Assoc. 74, 829–836, 1979.
7. Hand DJ. *Construction and Assessment of Classification Rules.* Chichester: John Wiley and Sons, 1997.
8. Hosmer DW and Lemeshow S. *Applied Logistic Regression*, Wiley, New-York, 1989.
9. Hosmer DW, Hosmer T, Le Cessie S, Lemeshow S. A Comparison of Goodness-of-fit Tests for the Logistic Regression Model. Statistics in Medicine 1997; 16:965–980.
10. de Keizer N. An Infrastructure for Quality Assessment in Intensive Care; Prognostic Models and Terminological Systems. PhD Thesis, 2000, University of Amsterdam.
11. Knaus W, Draper E, Wagner D, Zimmerman J. APACHE II: a Severity of Disease Classification System. Crit Care Med 1985; 13:818–829.
12. Kohavi R. Scaling Up the Accuracy of Naive-Bayes Classifiers: a Decision-Tree Hybrid. Proc. of the Second Int. Conference on Knowledge Discovery and Data Mining. 1996; 202–207.
13. Le Gall J, Lemeshow S, Saulnier F. A New Simplified Acute Physiology Score (SAPS-II) Based on a European/North American Multicenter Study. JAMA 1993; 270:2957–2963.
14. Long WJ. A Comparison of Logistic Regression to Decision-Tree Induction in a Medical Domain. Compt Bio Res 1993:74–97.
15. Lucas PJF and Abu-Hanna A. Prognostic Methods in Medicine (Lucas PJF and Abu-Hanna A. eds.). Special issue of Artificial Intelligence in Medicine. 1999; 15(2):105–119.
16. Marshall RJ. The use of Classification and Regression Trees in Clinical Epidemiology, Journal of Clinical Epidemiology 54 (6) (2001) pp. 603-609.
17. Miller ME and Hui SL. Validation Techniques for Logistic Regression Models. Statistics in Medicine, 1991, Vol 10, pp. 1213–1226.
18. Rowan K, Kerr J, Major E, McPherson K, Short A, Vessey M. Intensive Care Society's APACHE II Study in Britain and Ireland-II. BMJ 1993; 307:977–981.