

This Provisional PDF corresponds to the article as it appeared upon acceptance. Copyedited and fully formatted PDF and full text (HTML) versions will be made available soon.

Identification of high-risk subgroups in very elderly ICU patients

Critical Care 2007, **11**:R33 doi:10.1186/cc5716

Sophia E de Rooij (s.e.derooij@amc.nl)
Ameen Abu-Hanna (a.abu-hanna@amc.nl)
Marcel Levi (m.m.levi@amc.nl)
Evert de Jonge (e.dejonge@amc.nl)

ISSN 1364-8535

Article type Research

Submission date 16 November 2006

Acceptance date 8 March 2007

Publication date 8 March 2007

Article URL <http://ccforum.com/content/11/2/R33>

This peer-reviewed article was published immediately upon acceptance. It can be downloaded, printed and distributed freely for any purposes (see copyright notice below).

Articles in *Critical Care* are listed in PubMed and archived at PubMed Central.

For information about publishing your research in *Critical Care* go to

<http://ccforum.com/info/instructions/>

Identification of high-risk subgroups in very elderly ICU patients
*Sophia E. de Rooij*¹, *Ameen Abu-Hanna*², *Marcel Levi*³ and *Evert de Jonge*⁴

¹Head, Department of Geriatrics, Academic Medical Center, University of Amsterdam, Amsterdam

²Adjunct Head, Department of Medical Informatics, Academic Medical Center, University of Amsterdam, Amsterdam

³Professor and Head, Department of Internal Medicine, Cardiology and Pulmonary Disease, Academic Medical Center, University of Amsterdam, Amsterdam

⁴Adjunct Head Department of Intensive Care, Academic Medical Center, University of Amsterdam, Amsterdam

correspondence to:

Sophia E.J.A. de Rooij, MD

e-mail: s.e.derooij@amc.uva.nl

ABSTRACT

Introduction: Current prognostic models for ICU patients have not been specifically developed or validated in the very elderly. Aim of this study was to develop a prognostic model for ICU patients aged 80 years and older to predict in-hospital mortality using data obtained within 24 hours after ICU admission. Aside from good overall performance, the model should reliably identify specifically subgroups at very high risk of dying.

Methods: A total of 6867 consecutive patients aged 80 years and older from 21 Dutch ICUs were studied. Data necessary to calculate the Glasgow Coma Scale, APACHE II, SAPS II, MPMII scores, and ICU and hospital survival were recorded. Data were randomly divided in a developmental (n=4587) and a validation set (n=2289). Using recursive partitioning analysis a classification tree was developed predicting in-hospital mortality. This model was compared with the original SAPS II model and with the SAPS II model after recalibration for Dutch very elderly ICU patients.

Results: Overall performance measured by the area under the receiver operating characteristic curve and by the Brier-score was similar for the classification tree, the original SAPS II model and the recalibrated SAPS II model. The tree identified most patients with very high risk of mortality (9.2% of patients had risk > 80% vs. 8.9% for original SAPS II and 5.9% for recalibrated SAPS II). Using a cut-point at risk of 80%, the positive predictive value was 0.88 for the tree, 0.83 for original SAPS II and 0.87 for the recalibrated SAPS II.

Conclusion: Prognostic models with good overall performance may also reliably identify subgroups of very elderly ICU patients with a very high risk of dying before hospital discharge. The classification tree has the advantage of identifying the separate factors contributing to bad outcome and of using few variables. Up to 9.5% of patients were found to have a risk to die > 85%.

Introduction

The number of very elderly patients in the population has grown rapidly, and the next decades it will continue to increase even further [1]. At present, this ageing is both associated with an increased prevalence of co-morbidities and functional disabilities and with an increasing need for intensive care facilities. There is much uncertainty regarding which very elderly patients will benefit from ICU treatment and which subgroups may be identified to have very low or high risks of mortality.

Prognostic models, such as the Acute Physiology and Chronic Health Evaluation (APACHE) II or III [2,3], the Simplified Acute Physiology Score (SAPS) II [4], and the Mortality Probability Models (MPM) II [5] were developed to quantify the severity of illness and the likelihood of hospital survival for a general intensive care unit (ICU) population. These models should reliably predict the probability of mortality in all patients. However, little is known about the performance of these models in specific populations, such as very elderly patients. In addition, finding sub-groups of very elderly patients with a very high risk of dying may be important for several reasons. It identifies patients for whom better treatments are needed. At the same time, it may provide information to help patients and their caregivers to decide on intensive treatments that may be very burdensome. To decide on their willingness to receive intensive care treatment, very elderly patients want to know if they have a fair chance of surviving [6,7]. Also, identification of high-risk groups of patients may be useful for risk-stratification in scientific trials or for comparing outcome of different ICUs.

Aim of our study is to develop a prognostic model for very elderly ICU patients aged 80 years and older, that is able to reliably identify patients at very high risk of death before hospital discharge. In order to develop such a model we will use two statistical methods, namely a recalibrated SAPS II model based on logistic regression and the technique of recursive partitioning analysis (RPA). RPA is a nonparametric technique which iteratively subdivides a population in subgroups by creating mutually exclusive subsets according to a set of predictor variables. The process results in a classification tree.

Methods

Participants

We retrospectively studied 6867 consecutive patients aged 80 years and older admitted from January 1997 till December 2003 to the ICUs of 21 university, teaching, and non-teaching hospitals in the Netherlands. The data were obtained from the database of the Dutch National Intensive Care Evaluation (NICE) [8]. For the data analysis with recursive partitioning in this study we randomly divided the data into a developmental (n=4578) and a validation set (n=2289). The study was approved by the Medical Ethics Committee of our hospital, a tertiary university hospital.

Data collection

Data were collected as part of the NICE registry. For all patients, demographics, all data necessary to calculate the Glasgow Coma Scale, APACHE II [2] SAPS II [4], and MPMII [5] scores and ICU and hospital survival were recorded. To collect reliable data, NICE incorporates a framework of measures to improve data quality. Details concerning the quality of the data used in this study were published elsewhere [9].

Missing data

There were 7019 consecutive admissions in total. Records with missing values for admission type (N=142, of which 47 resulted in death) and SAPS-II scores (N=10) were excluded from the analysis resulting in 6867 admissions. GCS had 977 missing values, these were considered to be normal (value = 15) and have hence been imputed in the training and the validation sets. The number of missing values of other relevant variables varied from 0 to 10%: urine production within 24 hours (N=310); lowest bicarbonate (N=536); Urea (N=693); mechanical ventilation within 24 hours after admission (N=0); lowest systolic bloodpressure (N=214); lowest pH (N=670). The tree-fitting algorithm automatically handles missing values as described below.

Statistical Analysis

For continuous variables we use the t-distribution for calculating the 95% confidence intervals, and the Welch modification of the two-sample t-test for calculating the p-values for differences between means. This modification allows one not to assume equal variance in the survival and non-survival groups. We use Wilson's method for calculating the 95% confidence interval for proportions and binomial probabilities, such as mortality rate in the various patient subgroups and the positive predictive values (PPVs). The two-sided proportion test with Yates' continuity correction was used for testing differences between proportions, except for differences between PPVs for which bootstrapping (with 1000 bootstrap samples) was used because the patient groups partially overlap. Bootstrapping with 1000 bootstrap samples was also used to calculate the confidence interval of differences between Brier scores. The Hosmer-Lemeshow test with 10 degrees of freedom was used for testing model calibration.

In this study data were analyzed, among others, by means of recursive partitioning analysis (RPA) [10]. RPA forms an alternative to more standard model-based regression techniques for multivariable analyses. Contrary to such numeric-based-techniques, RPA results in a symbolic representation called a classification tree which can be easily interpreted as a collection of "If-Then rules", each with a condition part and a conclusion part. An example of a rule is "IF Glasgow Coma Scale > 6, AND patient is admitted to the ICU after planned surgery, AND the urine production over the first 24 hours > 1.25 L, THEN the risk to die before hospital discharge is 11.8%". The classification tree is obtained by finding the split—a variable and its value or cut-point value (e.g. Glasgow Coma Scale > 6)—that "best" partitions the whole group of patients into two subgroups. These subgroups, one fulfilling and one not fulfilling the condition in the split, appear graphically under a left and a right branch emanating from the group.

The term "best" refers to a partition resulting in the lowest entropy, in essence meaning that a probability of an event, such as survival status, differs most between the two subgroups. Next, each subgroup, is in turn itself further partitioned (hence the term "recursive partitioning" in RPA). This process is repeated until a stopping criterion is met. Each path from the root to a leaf node in the

tree corresponds to an If-Then rule where the conclusion part consists of the probability of the event in the leaf node.

When the tree algorithm finds the split that best partitions a group of observations, it also identifies "surrogate-splits" used to handle missing values. A surrogate-split partitions the observations in a very similar way to the original split (in terms of the "left and the right" subgroups). Suppose that the original split is "minimum bicarbonate < 22.6 $\mu\text{mol/L}$ " then for an observation missing the minimum bicarbonate value, the surrogate split "maximum bicarbonate < 25.3 $\mu\text{mol/L}$ " can be used to decide on whether the observation should go to the left or to the right branch. The surrogate-split mechanism is in effect a flexible way to impute a missing value depending on where it is encountered in the tree. The surrogate splitter contains information that is typically similar to what would be found in the primary splitter. Other products' approaches treat all records with missing values as if the records all had the same unknown value; with that approach all such 'missings' are assigned to the same bin. In CART, each record is processed using data specific to that record. This allows records with different data patterns to be handled differently, which results in a better characterization of the data.

In our study the root of the tree corresponds to the whole developmental sample and is associated with the prevalence (the a priori probability) of hospital mortality in the developmental set. Each variable is then assessed to determine which one discriminates most, in terms of information gain, between those who are discharged alive from hospital and those who did not survive hospital treatment.

This process is repeated on the new nodes creating as a result a tree structure. This process has been first allowed to completely over-grow the tree to overfit the data. Then the optimal tree size has been determined as the size that results in the minimum cross-validation error, as described below.

Then the original over-grown tree has been pruned back to the optimal size.

Cross validation is performed for increasing tree sizes (in essence this corresponds to the number of nodes in the tree). The cross-validation error is based on a ten-fold cross validation in which the developmental set is randomly split into ten mutually exclusive subsets. Nine sets are used to grow

a new tree of the given size, and the tenth is used to assess the accuracy of that tree in predicting the outcomes in this tenth subset. This process is repeated for each of the remaining nine sets to assess the performance, resulting in 10 error estimates. The cross validation error associated with the given tree size is the average value of the classification error of the ten trees of that size. The cross-validation error will usually first decrease with tree size, then reach a minimum which is associated with the optimal tree size, but then start increasing again due to overfitting.

The resulting pruned tree was then validated by measuring its predictive performance on the validation set which is not used in any way during the development of the tree. We used systematic sampling, including every third successive admission in the validation set.

We used the Rpart package, for recursive partitioning, and the glm function, for fitting logistic regression models, within the statistical environment S-PLUS (commercially available software, <http://www.lib.stat.cm.edu/S/rpart>) (MathSoft, Inc.: S-Plus: S-Plus Reference Manual, Seattle 1999).

The predictive ability of the tree was compared with the predicted mortality based on the original SAPS II score and with the predicted mortality based on the SAPS II model after first-level customization for a Dutch population of very elderly patients using the developmental database [11]. First level customization means refitting the model to obtain new coefficients without changing the score itself. Second level customization implies adapting each item of the score, this was not attempted here. A receiver operating characteristic (ROC) curve was generated for the logistic regression SAPS II models and the classification tree. The ROC curve represents a graphical display of sensitivity plotted against 1-specificity for all possible thresholds that can be used to predict hospital mortality. Estimates of the ROC-AUC and its standard error were obtained using DeLong et al's non-parametric approach [12]. The ROC-AUC measures the discriminative ability of a model. It is a non-strictly proper scoring rule: its maximum value can be obtained also when the predictions are not equal to the true probabilities. This is because it is not sensitive to the distance between the predicted probability and the true probability of an event, which is a measure of calibration. Therefore we also measured the Brier score (i.e. the mean of the squared errors of the

predictions) which is a proper scoring rule. We also performed a Hosmer Lemeshow test with 10 degrees of freedom.

Results

The overall mortality was $n=1433$ (31.3 %) of the developmental set and $n= 699$ (30.5 %) of the validation set (difference is not significant). The studied cohort had a mean age of 83.4 years (developmental set 83.3 years, validation set 83.5 years, not significant). Characteristics of patients are shown in Table 1 and Table 2.

Classification tree

The classification tree was obtained by binary recursive partitioning from the developmental data set and is presented in Figure 1. Note that a right branch always corresponds to the subgroup with the higher risk. Every patient fulfils the criteria of just one of the eleven mutually exclusive subgroups at the leaves of the trees. A leaf corresponds to a sub-group which is not further subdivided. The predicted likelihood to die before hospital discharge is given by the corresponding box. For example, all patients with a Glasgow Coma Scale score of more than six, admitted to the ICU after planned surgery and with urine production over the first 24 hours of more than 1.25 liter had a risk to die before hospital discharge of 11.8%. Likewise, all patients with a Glasgow Coma Scale score of less than seven had a risk of 89.2% (Figure 1). Of all 4578 patients in the developmental set, 435 (9.5%) had a risk higher than 85% and 484 patients (10.6%) had a risk higher than 70%.

Performance of classification tree compared with original SAPS II and recalibrated SAPS II model

Overall performance of the different models is shown in Table 3 and 4. Discrimination, i.e. the ability to distinguish between survivors and non-survivors is given by the area under the receiver operator curve (ROC-AUC), see also Figure 2. The accuracy of the predictions are given by the Brier score, i.e. the mean squared difference between the prediction and the actual outcome of all patients, lower Brier scores indicating higher accuracy. When tested on all patients in the validation set, the ROC-AUC was 0.77 for all three models (Table 3). Also, identical Brier scores were found for the three models. However, the Brier score is sensitive to calibration and it showed that the recalibrated SAPS II model was better than the original SAPS II model (95% CI 0.0016-0.0081). The 95% Confidence intervals for the classification tree vs the original SAPS II model resp. for the

classification tree vs the recalibrated SAPS II model were non significantly different ((-0.172 - 0.011) and (-0.0119-0.0139)).

We also performed a Hosmer-Lemeshow test within 10 degrees of freedom. It resulted in: an H-statistic of 64.3 (p-value < 0.00001) and a C statistic of 89 (p-value < 0.00001) for the original SAPS II. For the recalibrated SAPS II we found a H statistic of 9.5 (p-value = 0.49) and a C statistic of 21.6 (p-value of 0.02). The recalibrated SAPS II model is clearly much better.

To test the ability to identify high-risk patients, positive predictive value (PPV) was calculated for three risk levels corresponding to the following cut-points: 0.5, 0.7 and 0.8 (Tables 3 and 4). When tested on all patients in the validation set (Table 3) the recalibrated SAPS II model had the highest PPV for the lowest risk with cut-point 0.5 (non significant vs. tree; significant vs. original SAPS II). However, the classification tree had the highest PPV when patients were identified with a risk higher than 0.7 (PPV is 0.85, significant vs. original SAPS-II; non-significant vs. recalibrated SAPS II) and higher than 0.8 (PPV is 0.88, non-significant differences with original and recalibrated SAPS II). The classification tree, the original SAPS II and the recalibrated SAPS II model predicted a likelihood of more than 0.8 to die before hospital discharge in 210 (9.2%), 203 (8.9%) and 136 (5.9%) of 2289 patients in the validation database respectively.

Performance of the models in patients fulfilling the entry criteria of the SAPS II model

The original SAPS II model excludes many patients for estimation of the risk to die. To make a fair comparison, we also tested the three models in the patients of the validation set fulfilling the criteria of the SAPS II model (see Table 4). The most important group of patients that was excluded in this analysis corresponded to patients after cardiac surgery. Interestingly, overall performance, measured by the ROC-AUC and the Brier score were lower for all models compared with performance in the complete validation set. In this analysis the number of patients with an estimated risk higher than 0.8 and the PPV were largest for the classification tree model (no significant testing was attempted).

Combination of classification tree and recalibrated SAPS II

We tested the hypothesis that combining the classification tree with the recalibrated SAPS II model would lead to a higher positive predictive value for identifying high risk patients. In the complete validation database 112 (4.9%) of the patients had a predicted risk of mortality > 80% in both the classification tree and recalibrated SAPS II. Observed mortality in these patients was 105 (i.e. with a PPV of 94%).

Discussion

The results of this study show that it is possible to reliably identify a relatively high percentage of very elderly ICU patients with a very high risk to die before hospital discharge. Up to almost 10% of patients were shown to have a risk to die greater than 85%.

Although overall predictive performance in all very elderly patients was similar for the SAPS II model, the positive predictive value for high-risk subgroups was larger for the recalibrated SAPS II model and the classification tree, and the classification tree identified most patients at very high risk. This does not mean that classification tree based models are better than logistic regression based models. SAPS II was developed almost 20 years ago. We can not rule out that a new model based on logistic regression and specifically developed for very elderly ICU patients would have even better predictive power. This classification tree offers the advantage that the predictions are based on only eight variables, making it very easy to use the model. Furthermore, it clearly shows which parameters are related to a bad outcome. As low Glasgow Coma Scale scores appear to be most strongly related to death, this could prompt finding new treatment strategies for comatose very elderly patients. Another advantage of the classification tree is the symbolic representation which is more easy to interpret. RPA also automatically identifies the predictors, the cut-points and the interactions among all possibilities. Furthermore, missing values are systematically dealt with.

To our knowledge, we present the first validated prognostic model based on recursive partitioning, which is able to reliably identify high risk mortality groups, and which is developed and validated on a large group of patients. Our results are in line with other studies using a classification tree [21]. However, these studies were based upon populations of patients described merely by malignancies, or fitted on a small population, without performing validation on a separate validation set [22-24].

Identification of high-risk groups of patients may be important for several reasons. First, as already mentioned, it focuses attention on groups of patients for whom current treatments may be insufficient. This in itself could lead to improvement of care. Second, for some medical studies, enrolment of high-risk patients in clinical trials may provide the highest likelihood for finding a positive effect or facilitate investigating treatments with serious adverse side effects that are only acceptable if other treatments are not effective. Third, identification of high-risk subgroups may be

used for case mix correction when comparing outcome of very elderly patients in different ICUs. Fourth, it may be used for providing optimal information to patients, their relatives and to caregivers. Very elderly patients do not necessarily prefer intensive care treatment over palliative care aiming at comfort and pain relief. Interestingly, when presented hypothetical scenarios, patients state that they would decline intensive treatments if the likelihood of survival was very low [6].

There are some limitations to our findings. The classification tree based model is developed in a Dutch population of very elderly ICU patients. Before this model can be used in other countries, it should first be validated in an international population. Furthermore, the model is based on data from 1997 to 2003. As prognosis of ICU patients may change over time, repeated validation is necessary in the future, if data from the model is to be used to support individual decision making. Also, it is not known what the influence is of providing prognostic information of this kind to individual patients. In addition, when actually faced with a life threatening condition, do very elderly patients really prefer palliative care over life-sustaining treatments in those circumstances? Almost all very elderly patients with a very high risk to die are not able to express their preferences because they have decreased consciousness or are otherwise too ill. Consequently, decisions regarding life sustaining treatments are made by physicians and family members or other legal representatives [13,14]. Physicians are not always aware of the preferences of their seriously ill patients [17] and it is unknown to what extent end-of-life decisions by family members are influenced by the likelihood of survival [13,15]. For all these reasons, using prognostic models for individual decision making carries many dangers. They should not yet be used for this purpose and more research is clearly necessary. Nevertheless, adequate communication, good decision making, and respect for patients' autonomy are key determinants to patient and family satisfaction [16].

Prognostic models may also be used for triage decisions based on economical grounds. Intensive Care resources are limited and expensive [17]. It has been stated that, from an economical perspective, costs between \$50,000 and \$100,000 per year of life gained are acceptable in the United States [18,19]. One could argue that ICU treatments should only be given to patients with a fair chance of survival [17]. However, as consensus is lacking about the likelihood of survival that

is required to offer ICU treatment to (very elderly) patients who otherwise almost certainly die [20], we believe that current prognostic models should not be used for triage purposes.

In addition, other reliable parameters should be studied and added to the present and to newly to be developed prognostic models. For instance, the presence of cognitive or functional impairment may play an important role in clinical decision making in receiving life-sustaining treatment and therefore in prognosis [25]. But before adaptation of prognostic models is possible, more prospective studies need to be carried out to study the impact of pre-admission cognitive and functional impairment on short-term outcome like ICU and hospital mortality or long-term functional outcome, especially in the very elderly.

Conclusion

Our results show that present prognostic models may reliably identify subgroups of very elderly patients with a very high risk of dying before hospital discharge. We suggest that future research should focus on how prognostic models may support individual patients and their families in decision making to ensure that they receive care consistent with their preferences.

Key Messages

- prognostic models reliably identify subgroups of very elderly ICU patients with a high risk of dying before hospital discharge
- up to almost 10% of patients were shown to have a risk to die greater than 80%.
- overall predictive performance in all very elderly patients was similar for the SAPS II model, the recalibrated SAPS II and the classification tree model
- in the very elderly few predictors, such as used in the classification tree model resulted in similar performance to SAPS II model
- identification of high-risk groups of patients may be important for several reasons

Acknowledgements:

The Academic Medical Center funded the study with an unrestricted grant.

Competing Interests

All authors hereby confirmed that there are no conflicts of interest.

Authors' contributions:

SR and EJ are the principal investigators of the study; they designed the protocol and supervised its progress, EJ en AAH were involved in the acquisition of the data. AAH was responsible for the statistical analysis. SR, ML and EJ drafted the manuscript. All authors contributed to the interpretation of the data, revisions of the paper, and read and approved the final manuscript.

REFERENCES

- 1 U.S.Census Bureau 2000. Population Projections Of the United States by Age, Sex, Race, Hispanic Origin and Nativity: 1999–2100
- 2 Knaus WA, Draper EA, Wagner DP et al. APACHE II: a severity of disease classification system. *Crit Care Med* 1985; 13(10):818-829
- 3 Knaus WA, Wagner DP, Draper EA et al. The APACHE III prognostic system. Risk prediction of hospital mortality for critically ill hospitalized adults. *Chest* 1991; 100(6):1619-1636
- 4 Le Gall JR, Loirat P, Alperovitch A et al. A simplified acute physiology score for ICU patients. *Crit Care Med* 1984; 12(11):975-977
- 5 Lemeshow S, Teres D, Pastides H et al. A method for predicting survival and mortality of ICU patients using objectively derived weights. *Crit Care Med* 1985; 13(7):519-525.
- 6 Fried TR, Bradley EH, Towle VR et al. Understanding the treatment preferences of seriously ill patients. *N Engl J Med* 2002; 346(14):1061-1066
- 7 Murphy DJ, Burrows D, Santilli S et al. The influence of the probability of survival on patients' preferences regarding cardiopulmonary resuscitation. *N Engl J Med* 1994; 330(8):545-549
- 8 de Jonge E, Bosman RJ, van der Voort PH et al. [Intensive care medicine in the Netherlands, 1997-2001. I. Patient population and treatment outcome]. *Ned Tijdschr Geneesk* 2003; 147(21):1013-1017
- 9 Arts D, de Keizer N, Scheffer GJ et al. Quality of data collected for severity of illness scores in the Dutch National Intensive Care Evaluation (NICE) registry. *Intensive Care Med* 2002; 28(5):656-659
- 10 Breiman L, Olshen RA, Stone CJ et al. *Classification and Regression Trees*.1984
- 11 Zhu BP, Lemeshow S, Hosmer DW et al. Factors affecting the performance of the models in the Mortality Probability Model II system and strategies of customization: a simulation study. *Crit Care Med* 1996; 24(1):57-63
- 12 DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics* 1988; 44:837-845
- 13 Frank C, Heyland DK, Chen B et al. Determining resuscitation preferences of elderly inpatients: a review of the literature. *CMAJ* 2003; 169(8):795-799
- 14 Smucker WD, Houts RM, Danks JH et al. Modal preferences predict elderly patients' life-sustaining treatment choices as well as patients' chosen surrogates do. *Med Decis Making* 2000; 20(3):271-280
- 15 Fried TR, Gillick MR. The limits of proxy decision making: overtreatment. *Camb Q Healthc Ethics* 1995; 4(4):524-529

- 16 Heyland DK, Cook DJ, Rocker GM et al. Decision-making in the ICU: perspectives of the substitute decision-maker. *Intensive Care Med* 2003; 29(1):75-82
- 17 Teno JM, Fisher E, Hamel MB et al. Decision-making and outcomes of prolonged ICU stays in seriously ill patients. *J Am Geriatr Soc* 2000; 48(5 Suppl):S70-S74
- 18 Coughlin MT, Angus DC. Economic evaluation of new therapies in critical illness. *Crit Care Med* 2003; 31(1 Suppl):S7-16
- 19 Angus DC, Barnato AE, Linde-Zwirble WT et al. Use of intensive care at the end of life in the United States: an epidemiologic study. *Crit Care Med* 2004; 32(3):638-643
- 20 Azoulay E, Pochard F, Garrouste-Org et al. Decisions to forgo life-sustaining therapy in ICU patients independently predict hospital death. *Intensive Care Med* 2003; 29(11):1895-1901
- 21 Gilpin E, Olshen R, Henning H et al. Risk prediction after myocardial infarction. Comparison of three multivariate methodologies. *Cardiology* 1983; 70(2):73-84.
- 22 El Solh AA, Sikka P, Ramadan F. Outcome of older patients with severe pneumonia predicted by recursive partitioning. *J Am Geriatr Soc* 2001; 49(12):1614-1621
- 23 Williams SG, Millar JL, Dally MJ et al. What defines intermediate-risk prostate cancer? Variability in published prognostic models. *Int J Radiat Oncol Biol Phys* 2004; 58(1):11-18
- 24 Esteban A, Anzueto A, Frutos-Vivar F et al. Outcome of older patients receiving mechanical ventilation. *Intensive Care Med* 2004; 30(4):639-646
- 25 de Rooij SE, Abu-Hanna A, Levi M, de Jonge E. Factors that predict outcome of intensive care treatment in very elderly patients. A review. *Crit Care* 2005;9(4):307-314
- 26 Recursive partitioning in the health sciences. H. Zhang and B. Singer, Springer Verlag, New York, 1999. No. of pages: 230. ISBN 0-387-98671-5.

Table 1. Characteristics of patients surviving or not surviving until hospital discharge (developmental set)

<i>Variable</i>	<i>Survivors (n=3145)</i>	<i>Non survivors (n=1433)</i>	<i>p-value</i>
Age, mean \pm SD	83.2(3.3)	83.6(3.4)	<0.05
Male (%)	45.7	47.8	0.554
Temperature (max) degrees Celsius, mean \pm SD	37.5(0.82)	37.4 (1.37)	0.397
Heart rate, mean \pm SD	69.6(16.7)	67.4(29.1)	0.0107
Sodium mmol/L, mean \pm SD	137(4.5)	137(5.87)	0.00329
Potassium mmol/L, mean \pm SD	3.7(0.544)	3.78(0.739)	<0.001
Creatinine μ mol/L, mean \pm SD	104(71.8)	143(105)	<0.001
Bicarbonate mmol/L, mean \pm SD	21.4(4.02)	18.7(5.53)	<0.001
Albumin g/L, mean \pm SD	22.3(6.8)	20(7.71)	<0.001
pH, mean \pm SD	7.38(0.0837)	7.32(0.124)	<0.001
Urine output, L/24hrs, mean \pm SD	2.65(2.02)	1.75(1.78)	<0.001
Glasgow Coma Scale			
Score = 15	2950	1029	<0.05
Score < 15	195(6.2%)	404(28.2%)	
APACHE II score, mean \pm SD	16.7(5.46)	22.5(8.16)	<0.001
APACHE II predicted mortality, mean \pm SD	0.197(0.178)	0.432(0.262)	<0.001
SAPS II score, mean \pm SD	36.2(12.1)	54.2(20.4)	<0.001
SAPS II predicted mortality, mean \pm SD	0.219(0.193)	0.513(0.305)	<0.001
Cardiopulmonary resuscitation before admission(%)	2.7	15.2	< 0.001
Length of Stay at ICU, days; Median(IQR)	1.0(0.8-2.9)	1.9(0.7-5.7)	< 0.001
Length of Stay at hospital, days; Median (IQR)	14(9-25)	10(3.2-24)	0.04

Note: p<0.05 is significant

Table 2. Referring specialty (developmental set)

<i>Referring specialty</i>	<i>Survivors (n=3145)</i>	<i>Non survivors (n=1433)</i>
Internal medicine	343 (10.9)	307 (21.4)
Cardiology	192 (6.1)	187 (13.0)
Pulmonary diseases	91 (2.9)	69 (4.8)
Neurology	42(1.3)	43 (3.0)
Surgery	1249 (39.7)	627 (43.7)
Cardiothoracic surgery	889 (28.3)	97 (6.8)
Neurosurgery	40 (1.3)	16 (1.1)
Other	299 (9.5)	82 (5.7)
<i>Admissiontype</i>		
Medical	737(24%)	723(51%)
Unplanned surgery	497(16%)	323(23%)
Planned surgery	1861(60%)	359(26%)

Table 3. Performance of classification tree, original SAPS II and re-calibrated SAPS II in all patients in the independent validation set (n=2289).

	<i>Classification tree</i>	<i>SAPS II</i>	<i>Re-calibrated SAPS II</i>
ROC-AUC (std)	0.77 (0.01)	0.77 (0.01)	0.77 (0.01)
Brier-score	0.16	0.16	0.16

Threshold Positive Predictive Value (PPV)

	<i>Classification tree</i>	<i>SAPS II</i>	<i>Re-calibrated SAPS II</i>
0.5	0.69 (0.64-0.73) (died n=329, predicted to die n=480)	0.68 (0.64-0.72) (died n=340, predicted to die n=502)	0.71 (0.67-0.76) (died n= 305, predicted to die n=427)
0.7	0.85 (0.8-0.89) (died n=196, predicted to die n=230)	0.78 (0.73-0.82) (died n= 241, predicted to die n=309)	0.81 (0.76-0.86) (died n=176, predicted to die n=215)
0.8	0.88 (0.83-0.91) (died n=184, predicted to die n=210)	0.83 (0.77-0.87) (died n= 168, predicted to die n=203)	0.88 (0.81-0.92) (died n=119, predicted to die n=136)

PPV is Positive predictive value, ROC-AUC is area under receiver operator curve, * is significant.

Confidence intervals of differences between PPVs:
(an asterisk indicates statistical significance at the 0.05 level):

Classification tree vs SAPS II: the CI's for cut-off 0.5 respectively 0.7 and 0.8 are (-0.032, 0.047), (0.023, 0.121)*, (-0.006, 0.104)

Classification tree vs Recalibrated SAPS II: the CI's for cut-off 0.5 respectively 0.7 and 0.8 are (-0.07, 0.011), (-0.027, 0.086), (-0.057, 0.055)

Recalibrated SAPS II vs SAPS II: the CI's for cut-off 0.5 respectively 0.7 and 0.8 are (0.056, 0.016)*, (0.072, 0.004)*, (0.087, 0.003)*

Table 4. Performance of classification tree, original SAPS II and re-calibrated SAPS II in patients in the independent validation set that fulfil the entry criteria of SAPS II model (n=1594). PPV is Positive predictive value, ROC-AUC is area under receiver operating characteristic curve.

	<i>Classification tree</i>	<i>SAPS II</i>	<i>Re-calibrated SAPS II</i>
ROC-AUC (std)	0.72 (0.01)	0.75 (0.01)	0.75 (0.01)
Brier-score	0.19	0.18	0.18
Threshold PPV			
0.5	0.64 (348)	0.66 (365)	0.69 (308)
0.7	0.83 (144)	0.76 (218)	0.79 (144)
0.8	0.85 (129)	0.80 (137)	0.84 (87)

std is standard deviation

Figure 1. Classification tree to predict mortality before hospital discharge in patients 80 years and older admitted to the intensive care unit. Percentages represent the likelihood of in-hospital mortality for patients in the respective sub-group (with 95% CI). A sub-group with mortality risk of > 75 % is indicated by a double-framed box.

Figure 2. Receiver operator characteristic curves for the classification tree and the SAPS II model (logistic regression). The curves of the original SAPS II model and the re-calibrated SAPS II model are identical.

Figure 1.

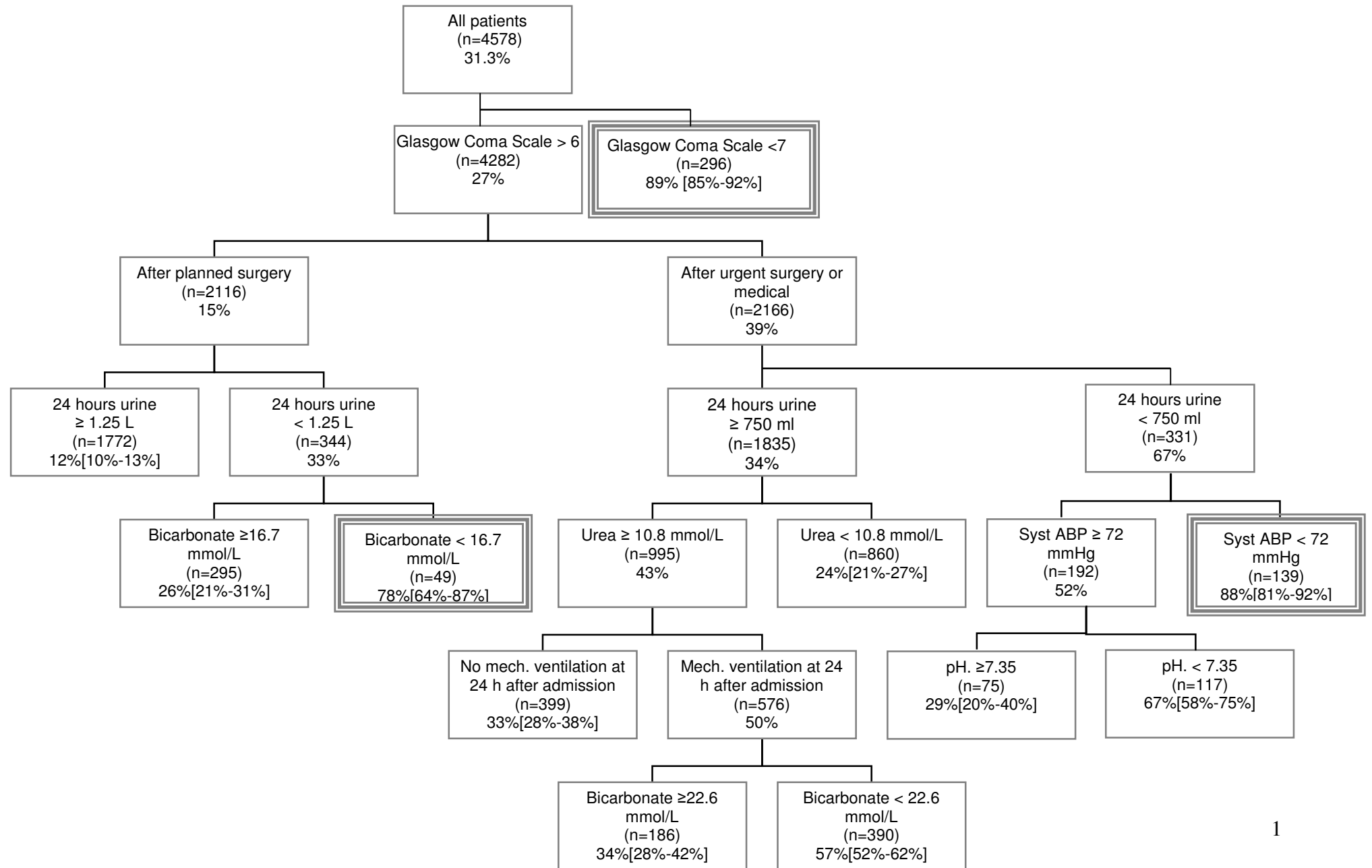


Figure 2.

