

Ashish Nimgaonkar
Dilip R. Karnad
S. Sudarshan
Lucila Ohno-Machado
Isaac Kohane

Prediction of mortality in an Indian intensive care unit

Comparison between APACHE II and artificial neural networks

Received: 3 March 2003
Accepted: 14 November 2003
Published online: 15 January 2004
© Springer-Verlag 2004

Electronic Supplementary Material
Supplementary material is available in the
online version of this article at
<http://dx.doi.org/10.1007/s00134-003-2105-4>

Part of this work was presented at the Sixth
Annual Critical Care Congress of the Indian
Society for Critical Care Medicine, Banga-
lore, India

A. Nimgaonkar (✉) · I. Kohane
Children's Hospital Informatics Program,
Ender's Building, 5th Floor,
320 Longwood Avenue,
Boston, Massachusetts, USA
e-mail: ashish@chip.org

A. Nimgaonkar · L. Ohno-Machado ·
I. Kohane
Division of Health Sciences
and Technology,
Harvard University and MIT,
Cambridge, Massachusetts, USA

D. R. Karnad
Medical Intensive Care Unit,
Department of Medicine,
KEM Hospital,
Parel, Mumbai, India

L. Ohno-Machado
Decision Systems Group,
Brigham and Women's Hospital,
Boston, Massachusetts, USA

A. Nimgaonkar
School of Biomedical Engineering,
Indian Institute of Technology,
Powai, Mumbai, India

S. Sudarshan
Department of Computer Science
and Engineering,
Indian Institute of Technology,
Powai, Mumbai, India

Abstract *Objective:* To compare hospital outcome prediction using an artificial neural network model, built on an Indian data set, with the APACHE II (Acute Physiology and Chronic Health Evaluation II) logistic regression model. *Design:* Analysis of a database containing prospectively collected data. *Setting:* Medical-neurological ICU of a university hospital in Mumbai, India. *Subjects:* Two thousand sixty-two consecutive admissions between 1996 and 1998. *Interventions:* None. *Measurements and results:* The 22 variables used to obtain day-1 APACHE II score and risk of death were recorded. Data from 1,962 patients were used to train the neural network using a back-propagation algorithm. Data from the remaining 1,000 patients were used for testing this model and comparing it with APACHE II. There were 337 deaths in these 1,000 patients; APACHE II predicted 246 deaths

while the neural network predicted 336 deaths. Calibration, assessed by the Hosmer-Lemeshow statistic, was better with the neural network ($\hat{H}=22.4$) than with APACHE II ($\hat{H}=123.5$) and so was discrimination (area under receiver operating characteristic curve =0.87 versus 0.77, $p=0.002$). Analysis of information gain due to each of the 22 variables revealed that the neural network could predict outcome using only 15 variables. A new model using these 15 variables predicted 335 deaths, had calibration ($\hat{H}=27.7$) and discrimination (area under receiver operating characteristic curve =0.88) which was comparable to the 22-variable model ($p=0.87$) and superior to the APACHE II equation ($p<0.001$). *Conclusion:* Artificial neural networks, trained on Indian patient data, used fewer variables and yet outperformed the APACHE II system in predicting hospital outcome.

Keywords Logistic regression · Outcome prediction · Prognostication · Severity scoring · Artificial intelligence · Information gain · Acute Physiology and Chronic Health Evaluation II · Goodness-of-fit

Introduction

Statistical techniques like multivariate regression have been used to develop outcome prediction models like APACHE, SAPS and MPM where age, pre-existing chronic diseases, clinical parameters, physiological derangements, surgical status and acute problems requiring ICU admission have been used to predict survival or death [1, 2, 3, 4, 5]. However, these statistical methods are constrained by the types of mathematical relationship between independent variables and outcomes that can be supported [6, 7]. Several models of artificial intelligence techniques have been used in the ICU [8, 9]. One such technique suited to predict mortality is the artificial neural network [8, 9] (See Appendix A for details on artificial neural networks). The commonest learning mechanism in artificial neural networks is the back-propagation algorithm, wherein the system predicts the outcome for each patient based on past experience (memory) and compares this with actual outcome. In this algorithm, an error is propagated backwards to update nodes and thus improve the prediction accuracy [6, 7].

Three previous studies have attempted to compare artificial neural network models with logistic regression models in small data sets in ICU settings [10, 11, 12]. However, these models have not been validated extensively. Our goal in this paper is to compare neural networks with an already validated and commonly used outcome prediction model. The Acute Physiology and Chronic Health Evaluation II (APACHE II) is a logistic regression model that is widely used and is considered the benchmark scoring system [13]. Moreover, besides in the US, it is applied in various countries including UK, France, Germany, Switzerland, Italy, Spain, Netherlands, Belgium, Saudi Arabia, Japan, New Zealand, Brazil and India [4, 5, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22].

The question remains whether a model such as APACHE II is better at predicting hospital outcome than a model derived from Indian patients treated in an Indian hospital. We therefore attempted to compare the predictive accuracy of artificial neural networks derived from Indian patients with the APACHE II scoring system.

Material and methods

The patients were aged more than 12 years and admitted to the 17-bedded Medical-neurological Intensive Care Unit of King Edward Memorial, Mumbai, India between January 1996 and May 1998. The King Edward Memorial Hospital is an 1,800-bed municipal hospital and tertiary referral center. In addition to patients with medical disorders, the medical intensive care unit also admits critically ill neurosurgical patients. The raw data used to obtain the APACHE II score were prospectively collected in all patients admitted to the ICU. The values recorded were the most abnormal/extreme physiological values during the first 24 h of ICU admission. The variables used in our study are given in Fig. 1. The hospital outcome (discharge or death) was also recorded. The

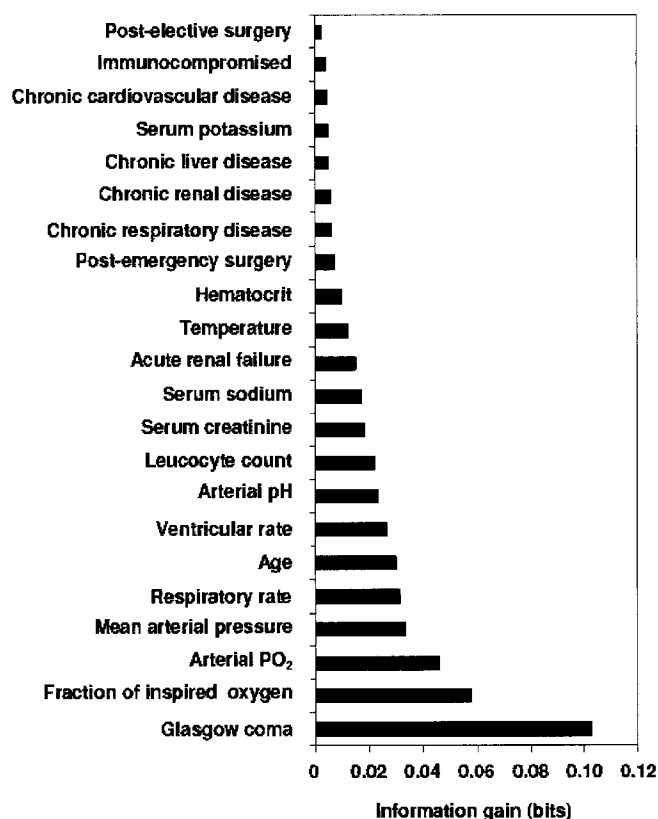


Fig. 1 Graphic representation of variables and their information gain using the training set of 1,962 intensive care unit patients

probability of hospital mortality, p , was derived from the APACHE II equation [6]:

$$\begin{aligned} \ln p/(1-p) = & -3.517 + (0.146 * \text{APACHE II score}) \\ & + (0.603 * \text{post-emergency operation status}) \\ & + \text{disease category coefficient} \end{aligned}$$

The artificial neural network architecture used was a feed-forward, back-propagation network with two hidden layers. A hidden layer is a layer of nodes between the input layer and output layer. Each hidden layer had 15 hidden units (nodes). Control for over-fitting was based on a 30% holdout subset of the training set. A learning rate of 0.01 was used. The hidden units weights were initialized with random numbers. Other parameters like momentum and number of iterations were optimized in an internal holdout set to obtain the best performance. The training was stopped when the predictive error reached a minimum on this set. All the artificial neural networks were implemented and tested using the DB2 Intelligent Miner software by IBM. The details of the technique are provided in Appendix A (Electronic Supplementary Material).

Two neural network models were developed, one with all the 22 variables and the other with 15 variables. These 15 variables were the variables with the highest information gain. Information gain is measured by calculation of entropy, which quantifies the effectiveness of a single variable or attribute in classifying the data that is used for training the artificial neural network [23]. The higher the information gain, the better is the variable in classifying the data into different categories. An example of this is discussed in Appendix B (Electronic Supplementary Material)

In the present study, 1,000 cases (of the total of 2,962 cases) were randomly selected and kept aside to serve as the test set (previously unseen cases in whom the artificial neural network would be used to see if it could correctly predict outcome). The remaining 1,962 cases were taken as the training set and a neural network model was developed based on these 1,962 cases. Initially all the 22 variables (attributes) that are used to predict mortality in the APACHE II model were taken to develop the network. While the APACHE II system uses PaO₂ for calculating the APACHE II score when the fraction of inspired oxygen (FIO₂) is less than 0.5 and the alveolar-arterial oxygen gradient when the FIO₂ is 0.5 or more, we used the PaO₂ and FIO₂ as input variables for the neural network. The APACHE II system considers the presence of any one chronic system disorder in assigning chronic health points, while we considered each of these five system disorders (namely chronic liver, renal, respiratory and cardiovascular dysfunction and immunocompromised state) separately in the artificial neural network model. Actual values were used for all variables except presence and absence of acute renal failure and the chronic disease variables.

A subsequent review of the entropy or information gain of the 22 variables revealed that 7 variables contributed very little to outcome prediction. Therefore only 15 variables with highest entropy were selected for building the next neural network model. All three models, namely the APACHE II, artificial neural network model with 22 input variables (ANN₂₂) and artificial neural network model with 15 input variables (ANN₁₅) were then used to predict the outcome in the validation set of 1,000 patients and the accuracy of these three methods in outcome prediction was compared.

As a statistical method, the Hosmer-Lemeshow statistic (\hat{H}) was used to study the calibration (the accuracy of the system in predicting group outcomes over the entire range of outcome risks) [24]. The Hosmer-Lemeshow statistic (\hat{H}) was tested on a chi-square distribution with eight degrees of freedom for the developmental set and ten degrees of freedom for the validation set [25, 26]. The area under the receiver operating characteristic curve was used to assess discrimination (the ability of the system to distinguish between individual patients who lived and those who died). The calculation for the area under the receiver operating characteristic curve was made using the method described by Hanley and McNeil [27, 28].

Results

We studied 2,962 ICU patients (Table 1). Of these cases, 1,000 cases were randomly selected and kept aside as the test cases. Using data from the remaining 1,962 cases, the ANN₂₂ was built taking 22 variables. The entropy (information gain) for all the variables are given in Fig. 1. From these data, it can be inferred that Glasgow Coma Score is the factor that would best predict prognosis, followed by the FIO₂, PO₂ and so on. As a second stage we then eliminated the seven variables with least information gain and included only the top 15 variables according to their information gain. Another artificial neural network model was built with these 15 variables using data from the same 1,962 patients.

All three models, the APACHE II, ANN₂₂ and ANN₁₅, were then used to predict the outcome in the test set of 1,000 patients and the accuracy of these three methods in outcome prediction were compared (Table 2). Using the Hosmer-Lemeshow statistic (\hat{H}) for evaluating the calibration of the

Table 1 Characteristics of 2,962 patients admitted to the Medical ICU of the King Edward Memorial Hospital, Mumbai, India

Variables	Number of patients (%)
Age in years (mean \pm SD)	37.5 \pm 16.1
Sex	
Males	1745 (58.9)
Females	1217 (41.1)
APACHE II score (mean \pm SD)	17.8 \pm 10.5
Operative status	
Non-operative	2378 (80.3)
Post-elective surgery	357 (11.0)
Post-emergency surgery	257 (8.7)
Chronic organ insufficiency or immunocompromised state*	348 (11.8)
Liver	65 (2.2)
Renal	62 (2.1)
Cardiovascular	128 (4.3)
Respiratory	81 (2.7)
Immunocompromised	81 (2.7)
Ten most common diagnoses	
Severe falciparum malaria	264 (8.9)
Poisoning	172 (5.8)
Diabetic ketoacidosis	156 (5.3)
Tetanus	150 (5.1)
Obstetric disorders	149 (5.0)
Postoperative intracranial neoplasm	142 (4.8)
Subarachnoid hemorrhage	117 (4.0)
Guillain Barre syndrome	100 (3.4)
Severe bacterial sepsis	85 (2.9)
Leptospirosis	84 (2.8)
Length of ICU stay in days (mean \pm SD)	5.8 \pm 8.3

* Criteria defined in the APACHE II system were used to diagnose chronic organ insufficiency or immunocompromised state [1]

three systems, the value of \hat{H} for the ANN₂₂ was 22.4, for the ANN₁₅ it was 27.7 and for APACHE II it was 123.5 (Fig. 2). Even though the values of \hat{H} for the two artificial neural network models were less than that for the APACHE II model, all the three models displayed significantly poor fit ($p < 0.05$) on the validation database when the Hosmer-Lemeshow statistic (\hat{H}) was tested on a chi-square distribution with ten degrees of freedom. The area under the receiver operating characteristic curve for APACHE II was 0.77. This was significantly less compared to ANN₂₂, which was 0.87 ($p < 0.002$), and ANN₁₅, which was 0.88 ($p < 0.001$), suggesting that the artificial neural network models were able to distinguish between survivors and non-survivors more reliably than APACHE II.

On the developmental set, the ANN₂₂ had an area under the receiver operating characteristic curve of 0.887 and Hosmer-Lemeshow statistic (\hat{H}) of 39.4. For the ANN₁₅, the model on the developmental set had an area under the receiver operating characteristic curve of 0.884 and Hosmer-Lemeshow statistic (\hat{H}) of 52.1. For the APACHE II, the area under receiver operating characteristic curve on the development set was 0.767 and Hosmer-Lemeshow statistic (\hat{H}) was 248.7.

Table 2 Data of 1,000 patients (test set) grouped in deciles of risk of death, along with the predicted and actual deaths in each category using all the three prediction models

Risk of death	APACHE II		ANN ₂₂ (all 22 variables)			ANN ₁₅ (15 variables)			
	Total	Deaths		Total	Deaths		Total	Deaths	
		Obs	Pred		Obs	Pred		Obs	Pred
0-10	361	44	12.94	215	20	12.71	221	26	12.60
10-20	210	54	28.63	190	32	25.91	179	22	24.28
20-30	104	41	24.90	141	40	35.34	151	34	36.52
30-40	90	35	30.92	100	28	35.19	101	40	34.99
40-50	73	39	32.52	86	40	38.64	88	38	39.38
50-60	38	24	21.14	57	31	30.76	52	22	28.08
60-70	41	28	26.84	93	48	59.92	73	40	46.90
70-80	36	31	27.19	50	33	37.61	55	48	40.77
80-90	36	31	30.41	41	39	34.40	44	34	37.27
90-100	11	10	10.41	27	26	25.31	36	33	33.84
Total	1000	337	245.9	1000	337	336	1000	337	335
Lemeshow-Hosmer statistic	$\chi^2=123.5$ df=10			$\chi^2=22.4$ df=10			$\chi^2=27.7$ df=10		

Obs observed, Pred predicted

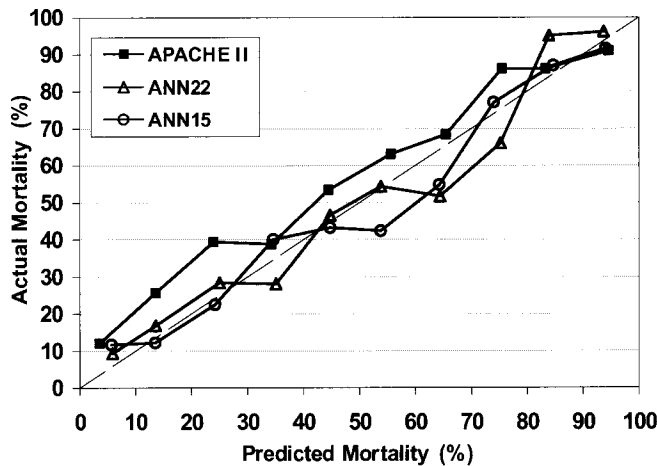


Fig. 2 Calibration curves for the APACHE II system and the two artificial neural network models in 1,000 Indian ICU patients (test set). ANN₂₂ is the artificial neural network model which uses 22 variables used by the APACHE II system and the ANN₁₅ model uses only those 15 variables from the ANN₂₂ model that had the highest information gain

Discussion

We found that the performance of the two models of neural networks was significantly superior to that of the APACHE II model when applied to the given data set. The area under the receiver operating characteristic curve for the artificial neural networks models showed better discriminative capabilities (accuracy of predicting whether an individual patient would survive or die) as compared to APACHE II model. However, the Hosmer-Lemeshow statistic (\hat{H}) showed that overall goodness-of-fit with the artificial neural network models was comparable to the APACHE II model. This statistical test is used to compare prediction of group outcomes over the entire range of predicted outcomes [4, 24].

While some studies have shown neural networks superior to regression models for some medical problems [10, 11, 29], others have shown no significant differences between the regression and neural network models [12]. Three previous papers have compared the artificial neural networks with the logistic regression model to predict ICU outcome [10, 11, 12]. All three studies collected clinical data needed for the APACHE score, but they did not use the APACHE equation to predict outcome. Instead, they used this raw data to develop their own logistic regression equation to predict survival or death. All three found that the predictive performance of artificial neural networks was comparable to the derived logistic regression, but was not superior. However, the logistic regression equations developed by these authors were trained on relatively small data sets and have not been validated elsewhere [10, 11, 12]. On the other hand, the original APACHE II equation for predicting outcome was derived from a very large cohort of American ICU patients and it has been extensively used in inter-hospital and international comparisons of ICU outcome. APACHE II is currently the most commonly used prognostic model in Indian ICUs. The performance of artificial neural network models has not been compared with that of a currently accepted and internationally validated logistic regression model such as APACHE. We compared our artificial neural network models with APACHE II for outcome prediction.

Two previous studies have attempted this comparison. Wong and Young [29] compared prediction by APACHE II with that of an artificial neural network in patients admitted to ICUs in the UK. Their study showed that there was no significant difference between the two approaches for predicting survival. Additionally, they showed that an artificial neural network could predict outcome without using the disease category coefficient, which is required by the APACHE II equation [29]. Frize et al. [8] studied 1,491 patients admitted to a Canadian ICU. Data from two-thirds of these patients were used for training the neural network

and the remaining one-third for validation. These authors, too, found similar accuracy in predicting outcome using the artificial neural network model and APACHE II. However, the artificial neural network could predict outcome using only six variables used by APACHE II.

In another study from the UK, predicted outcomes of patients with trauma using the Trauma and Injury Severity Score (TRISS) model and neural networks were compared. In this study, TRISS had a better discrimination while the artificial neural network had better calibration/goodness-of-fit [30]. The authors observed that the TRISS model, which assumes a linear relationship between the predictor variables and outcome, had better discrimination than the neural network. However, the neural network was able to deal with non-linear variables better and had better calibration than the TRISS model. As in our study, these authors found that the artificial neural network could predict outcome using fewer variables [30].

There are several possible explanations for the apparent superiority of the artificial neural network models over the APACHE II model in our patients. Although the APACHE II variables have points assigned by experts, the final APACHE II mortality prediction equation is derived using a logistic regression approach, which assumes a semi-linear relationship between the predictor variables and outcome. Neural networks are good at building non-linear models and, therefore, may offer at least a theoretical advantage. More importantly, the APACHE II model has been primarily derived based on a western cohort that is not representative of the ICU patient population of the Indian subcontinent [22, 31]. There is a significant difference between the case-mix in Indian and American ICUs (Table 1); this may also have affected the accuracy of the APACHE II model. Indian ICU patients also differ from American and European ICU patients with respect to other factors which influence outcome, including lead-time bias and differences in organization and utilization of health care resources [22, 31]; one such organizational difference was inclusion of patients aged 13–18 years, who are normally treated in “adult” medical facilities in India.

Standards of care and the availability of human and material resources are also likely to be different in American and Indian ICUs. Hence the APACHE II score, with its existing weights, showed good discrimination, but poor calibration in Indian patients. On the other hand, our artificial neural network models were trained using Indian patient data and may, therefore, have outperformed the APACHE II system on this basis alone. This could also explain why Wong and Young did not find artificial neural networks to be superior to APACHE II in patients from the UK [29], nor did Frize et al. [8] in Canadian patients. The demographic characteristics and case-mix of their patients may have been relatively similar to the original APACHE II cohort. Hence the APACHE II system may have accurately predicted outcome in their cohort such that little additional benefit could have been accrued from artificial neural networks.

Another significant observation in our study was that some of the variables used by the logistic regression model were redundant and did not contribute to improving the accuracy of prediction and, hence, could be eliminated from the model-building process. Least information gain was obtained from the chronic health evaluation variables, and the variables that were most useful were ones that assessed acute physiological status. Moreover, the artificial neural network models remained fairly accurate despite the exclusion of the diagnostic disease category for which a patient was admitted to the ICU. Wong and Young were able to eliminate some variables from the APACHE II model without loss of accuracy. Clermont et al. [12] also found outcome prediction was good, even after the exclusion of some variables like admission diagnosis and location prior to ICU admission. Frize et al. [8] could predict outcome with artificial neural networks using only six variables used by the APACHE II system. It therefore appears that, while artificial neural network models may be as good if not better than APACHE II for outcome prediction, their greatest strength may be in their ability to do so with fewer variables.

Variable selection is important in model development because the presence of too many inputs in a prediction system can decrease its performance by leading to model over-fitting and by adding more complexity to the model. For k number of variables, there are 2^k models that need to be evaluated to find the most optimal model. This problem is computationally intractable. For example, with 10 variables, the number of possible models is 1,024 and when the variables grow to 20, the number of models becomes 1,048,576. Variable selection methods that have been used in predictive model development include backward selection, forward selection and stepwise regression in logistic regression models, automatic relevance determination in Bayesian neural networks, rough sets and genetic algorithms [32, 33, 34]. We used the measure of entropy for variable selection. The application of entropy variable selection in forward, backward and stepwise procedures is possible and actually implemented in statistical packages such as SAS by selection using the Akaike information criterion (AIC) [35].

In conclusion, our study shows that, in the Indian ICU cohort, artificial neural network models built from a single-center Indian cohort showed better discrimination than the APACHE II model in predicting hospital outcome; the calibration (goodness-of-fit) of both models was poor. Our study and previous reports have shown that it is possible to predict outcome reliably using fewer variables than those needed in APACHE II. This could be an important benefit as the use of fewer variables will decrease the time, effort and cost involved in the collection of prognostic data [8, 10, 29]. Studies in larger, more heterogeneous ICU patient populations are needed to confirm our observation.

References

1. Knaus WA, Draper EA, Wagner DP, Zimmerman JE (1985) APACHE II: a severity of disease classification system. *Crit Care Med* 13:818–829
2. Knaus WA, Wagner DP, Draper EA, Zimmerman JE, Bergner M, Bastos PG, Sirio CA, Murphy DJ, Lotring T, Damiano (1991) The APACHE III prognostic system. Risk prediction of hospital mortality for critically ill hospitalized adults. *Chest* 100:1619–1636
3. Le Gall JR, Lemeshow S, Saulnier F (1993) A new Simplified Acute Physiology Score (SAPS II) based on a European/North American multicenter study. *JAMA* 270:2957–2963
4. Lemeshow S, Le Gall JR (1994) Modeling the severity of illness of ICU patients. A systems update. *JAMA* 272:1049–1055
5. Lemeshow S, Teres D, Klar J, Avrunin JS, Gehlbach SH, Rapoport J (1993) Mortality Probability Models (MPM II) based on an international cohort of intensive care unit patients. *JAMA* 270:2478–2486
6. Cross SS, Harrison RF, Kennedy RL (1995) Introduction to neural networks. *Lancet* 346:1075–1079
7. Hinton GE (1992) How neural networks learn from experience. *Sci Am* 267:144–151
8. Frize M, Ennett CM, Stevenson M, Trigg HC (2001) Clinical decision support systems for intensive care units: using artificial neural networks. *Med Eng Phys* 23:217–225
9. Hanson CW 3rd, Marshall BE (2001) Artificial intelligence applications in the intensive care unit. *Crit Care Med* 29:427–435
10. Dybowski R, Weller P, Chang R, Gant V (1996) Prediction of outcome in critically ill patients using artificial neural network synthesised by genetic algorithm. *Lancet* 347:1146–1150
11. Doig GS, Inman KJ, Sibbald WJ, Martin CM, Robertson JM (1993) Modeling mortality in the intensive care unit: comparing the performance of a back-propagation, associative-learning neural network with multivariate logistic regression. *Proc Annu Symp Comput Appl Med Care* 361–365
12. Clermont G, Angus DC, DiRusso SM, Griffin M, Linde-Zwirble WT (2001) Predicting hospital mortality for patients in the intensive care unit: a comparison of artificial neural networks with logistic regression models. *Crit Care Med* 29:291–296
13. Angus DC, Sirio CA, Clermont G, Bion J (1997) International comparisons of critical care outcome and resource consumption. *Crit Care Clin* 13:389–407
14. Markgraf R, Deutschinoff G, Pientka L, Scholten T (2000) Comparison of acute physiology and chronic health evaluations II and III and simplified acute physiology score II: a prospective cohort study evaluating these methods to predict outcome in a German interdisciplinary intensive care unit. *Crit Care Med* 28:26–33
15. Von Bierbrauer A, Riedel S, Cassel W, von Wichert P (1998) Validation of the acute physiology and chronic health evaluation (APACHE) III scoring system and comparison with APACHE II in German intensive care units. *Anaesthesist* 47:30–38
16. Jacobs S, Chang RW, Lee B (1988) Audit of intensive care: a 30-month experience using the APACHE II severity of disease classification system. *Intensive Care Med* 14:567–574
17. Arabi Y, Haddad S, Goraj R, Al-Shimemeri A, Al-Malik S (2002) Assessment of performance of four mortality prediction systems in a Saudi Arabian intensive care unit. *Crit Care* 6:166–174
18. Sirio CA, Tajimi K, Tase C, Knaus WA, Wagner DP, Hirasawa H, Sakanishi N, Katsuya H, Taenaka N (1992) An initial comparison of intensive care in Japan and the United States. *Crit Care Med* 20:1207–1215
19. Abu-Zidan FM, Plank LD, Windsor JA (2002) Proteolysis in severe sepsis is related to oxidation of plasma protein. *Eur J Surg* 168:119–123
20. Cavalcante NJ, Sandeville ML, Medeiros EA (2001) Incidence of and risk factors for nosocomial pneumonia in patients with tetanus. *Clin Infect Dis* 33:1842–1846
21. Shukla VK, Ojha AK, Pandey M, Pandey BL (2001) Pentoxifylline in perforated peritonitis: results of a randomised, placebo controlled trial. *Eur J Surg* 167:622–624
22. Parikh CR, Karnad DR (1999) Quality, cost and outcome of intensive care in a public hospital in Bombay, India. *Crit Care Med* 27:1754–1759
23. Tom M (1997) Decision tree learning. *Machine learning*. McGraw-Hill, New York, pp 55–60
24. Lemeshow S, Hosmer DW Jr (1982) A review of goodness of fit statistics for use in the development of logistic regression models. *Am J Epidemiol* 115:92–106
25. Hosmer DW, Lemeshow S (1980) A goodness-of-fit test for the multiple logistic regression model. *Comm in Stat A10:1043–1069*
26. Hosmer DW, Lemeshow S (1989) Applied logistic regression. Wiley, New York
27. Hanley JA, McNeil BJ (1982) The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* 143:29–36
28. Hanley JA, McNeil BJ (1983) A method of comparing the areas under receiver operating characteristic curves derived from the same cases. *Radiology* 148:839–843
29. Wong LS, Young JD (1999) A comparison of ICU mortality prediction using the APACHE II scoring system and artificial neural networks. *Anaesthesia* 54:1048–1054
30. Becalick DC, Coats TJ (2001) Comparison of artificial intelligence techniques with UKTRISS for estimating probability of survival after trauma. *UK Trauma and Injury Severity Score. J Trauma* 51:123–133
31. Sachdeva RC, Guntupalli KK (1999) International comparisons of outcomes in intensive care units. *Crit Care Med* 27:2032–2033
32. Atienza F, Martinez-Alzamora N, De Velasco JA, Dreiseitl S, Ohno-Machado (2000) Risk stratification in heart failure using artificial neural networks. *Proc AMIA Symp* pp 32–36
33. Dreiseitl S, Ohno-Machado L, Vinterbo S (1999) Evaluating variable selection methods for diagnosis of myocardial infarction. *Proc AMIA Symp* pp 246–250
34. Vinterbo S, Ohno-Machado L (2000) A genetic algorithm approach to multi-disorder diagnosis. *Artif Intell Med* 18:117–132
35. Akaike H (1974) A new look at the statistical model identification. *IEEE Trans Autom Contr* 19:716–723