

LIVER

Artificial neural network is superior to MELD in predicting mortality of patients with end-stage liver disease

A Cucchetti, M Vivarelli, N D Heaton, S Phillips, F Piscaglia, L Bolondi, G La Barba, M R Foxton, M Rela, J O'Grady, A D Pinna

Gut 2007;56:253–258. doi: 10.1136/gut.2005.084434

See end of article for authors' affiliations

Correspondence to:
Dr M Vivarelli, Dipartimento di Discipline Chirurgiche, Rianimatorie e dei Trapianti Policlinico S Orsola Via Massarenti, 9 40138 Bologna, Italy; vivarelli@aosp.bo.it

Revised 4 April 2006
Accepted 14 June 2006
Published Online First
29 June 2006

Background: Despite its accuracy, the model for end-stage liver disease (MELD), currently adopted to determine the prognosis of patients with liver cirrhosis, guide referral to transplant programmes and prioritise the allocation of donor organs, fails to predict mortality in a considerable proportion of patients.

Aims: To evaluate the possibility to better predict 3-month liver disease-related mortality of patients awaiting liver transplantation using an artificial neural network (ANN).

Patients and methods: The ANN was constructed using data from 251 consecutive people with cirrhosis listed for liver transplantation at the Liver Transplant Unit, Bologna, Italy. The ANN was trained to predict 3-month survival on 188 patients, tested on the remaining 63 (internal validation group) unknown by the system and finally on 137 patients listed for liver transplantation at the King's College Hospital, London, UK (external cohort). Predictions of survival obtained with ANN and MELD on the same datasets were compared using areas under receiver-operating characteristic (ROC) curves (AUC).

Results: The ANN performed significantly better than MELD both in the internal validation group (AUC = 0.95 v 0.85; $p = 0.032$) and in the external cohort (AUC = 0.96 v 0.86; $p = 0.044$).

Conclusions: The ANN measured the mortality risk of patients with cirrhosis more accurately than MELD and could better prioritise liver transplant candidates, thus reducing mortality in the waiting list.

Liver transplantation is an accepted treatment for patients with end-stage chronic liver disease. Owing to the imbalance between the potential number of recipients and the donor shortage, it is of paramount importance to accurately predict the prognosis of patients with liver cirrhosis to establish the correct timing of referral to a liver transplant programme on one hand and prioritise the allocation of organs to the most ill patient already on the waiting list according to guidance issued by the Department of Health and Human Services in 1998¹ on the other.

The model for end-stage liver disease (MELD) score,² originally introduced to measure the mortality risk in patients with end-stage liver disease undergoing transjugular intrahepatic portosystemic shunt, has been subsequently adopted for use as a disease severity index to determine referral to transplant programmes and organ allocation priorities. Although it is the most accurate scoring system currently available, it still fails to predict the mortality of a considerable proportion of patients.^{3,4}

The ability of MELD to predict mortality, expressed as the c statistic, ranges from 0.78 to 0.87.³ The model is based on serum bilirubin, creatinine and international normalised ratio (INR); the effect of other prognostic information, namely, the presence of refractory ascites, variceal bleeding, spontaneous bacterial peritonitis and encephalopathy have failed to provide further prognostic information. Little change has been observed in the ability of the MELD score to predict 3-month mortality by adding these other parameters.^{3,5}

A possible explanation is that patients with cirrhosis represent a biological system where the relationship between variables that determine the prognosis is complex, multi-dimensional and non-linear, so that it is difficult to distinguish between classes with the conventional linear discriminant analysis. When two classes are separated by a non-linear boundary, artificial neural networks (ANNs) perform better than conventional discriminant analysis.^{6,7}

ANN use computer technology to model systems that structurally and functionally recall biological neural networks: they consist of a set of highly interconnected processing units (neurons) linked with weighted connections, and include an input layer, an output layer and one or more hidden layers. The input layer is formed by the different data available for the analysis (eg, various laboratory tests) and the output layer is formed by the different outcomes.

One of the basic characteristics of the ANN is that it can learn through examples and therefore establish the weight to be given to any input. In detail, learning occurs through exposure to paired input–output data (training). The ANN learns to associate each of the inputs with the corresponding output, by modifying the weight of the connections between its neurons. Once an input is applied as a stimulus to the first layer of neurons, it is propagated through each upper layer until an output is produced. This output pattern is then compared with the desired output and an error signal is generated. The error signal is then transmitted backwards across the net and the weight of the connections between neurons is updated to decrease the overall error of the network. As learning proceeds, the difference between the ANN output and the desired output decreases to a minimum. On the basis of the knowledge accumulated during the training, the ANN can assign outputs to new input data not used in the learning process: after training the ANN can therefore identify patterns or make predictions on datasets never seen before.^{6–8}

In this study, we assessed the ability of the ANN to predict the 3-month mortality in patients with end-stage liver disease, using clinical data of patients with cirrhosis recorded at the

Abbreviations: ALT, alanine transaminase; ANN, artificial neural network; AST, aspartate transaminase; AUC, area under the curve; INR, international normalised ratio; MELD, model for end-stage liver disease; ROC, receiver-operating characteristic; UNOS, United Network for Organ Sharing

time of entering the waiting list for liver transplantation, and compared the prognostic performance of the ANN with that of the MELD.

PATIENTS AND METHODS

To train the ANN to predict the 3-month mortality of patients awaiting liver transplantation and to validate the system, a population of patients listed for liver transplantation at the Department of Surgery and Transplantation, University of Bologna, Italy, was retrospectively reviewed (internal cohort).

To obtain an external validation of the ANN, this was retrospectively tested on a population of patients listed for liver transplantation at the King's College Hospital, London, UK (external cohort).

In both the internal and external cohorts, the ANN was compared with the MELD scoring system in terms of accuracy in predicting 3-month mortality.

Internal cohort

From January 1999 to April 2003, 417 adults with end-stage chronic liver disease were listed for liver transplantation at the Department of Surgery and Transplantation, University of Bologna. As this study aimed to assess the ability of the neural network to predict 3-month mortality, the analysis did not include 155 patients whose waiting time until transplantation was <3 months. To avoid any possible bias related to a previous liver transplant that correlates with a higher patient mortality,⁹ 11 patients awaiting liver retransplantation, owing to chronic rejection ($n=3$) and recurrent hepatitis ($n=9$), were also excluded from the analysis. The remaining 251 patients represented the study group (internal cohort).

Until April 2003, liver grafts were allocated to potential recipients in accordance with the United Network for Organ Sharing (UNOS) status (status 2A, patients with chronic liver disease hospitalised in the intensive care unit with a life expectancy <7 days without transplantation; status 2B, patients hospitalised requiring continuous medical care but not in the intensive care unit; and status 3, patients requiring medical care but not hospitalised), on Child–Turcotte–Pugh score, ABO blood-type compatibility and overall waiting time.¹⁰ Patients who received liver transplantation after a waiting time of >3 months were considered alive at this time point. Patients who were dropped from the waiting list for reasons not related to liver disease or who died later than 3 months of waiting time were also considered alive at 3 months.

Table 1 reports the clinical data that were collected for each of the 251 patients. Aetiology of liver disease and diagnosis of cirrhosis were based on virological markers (hepatitis B and C serology), abdominal imaging (Doppler ultrasound and computer tomography) and assessment of biochemical liver function tests (serum bilirubin, albumin, transaminase and INR). Diagnosis of primary cholestatic disease (primary biliary cirrhosis and primary sclerosing cholangitis) was always confirmed with a liver biopsy. The MELD score was calculated according to the formula proposed by UNOS as follows:

$$0.957 \times \text{Ln}(\text{creatinine mg/dl}) + 0.378 \times \text{Ln}(\text{bilirubin mg/dl}) + 1.120 \times \text{Ln}(\text{INR}) + 0.643$$

laboratory values <1.0 were set to 1.0; the maximum serum creatinine considered was 4.0 mg/dl; the score was finally multiplied by 10 and rounded to the nearest whole number.

During the 3-month follow-up period 38 (15%) patients died. Of the 213 alive at 3 months, 203 patients then underwent transplantation, 3 were excluded for reasons not related to liver disease and 7 died after this period.

External cohort

In all, 371 adults with chronic liver disease listed for liver transplantation in the same period at the Liver Transplant Surgical Service, King's College Hospital, London, UK, were retrospectively evaluated. Inclusion criteria were the same as in the internal cohort: 137 of these patients were eligible for the study; table 1 reports clinical data available for this group; 226 patients whose waiting time to transplantation was <3 months and 8 patients awaiting liver retransplantation were excluded from the analysis. During the 3-month follow-up period 17 (12.4%) patients died. All of the 120 patients alive at 3 months underwent transplantation.

DEVELOPMENT OF THE ARTIFICIAL NEURAL NETWORK

The neural network was developed using NeuroSolution V.4 (Neurodimension, Florida, USA). Neurosolution is a user-friendly software with a Microsoft Excel add-in that simplifies the construction of the ANN. A licensed copy was bought but a free evaluation edition can be downloaded at <http://www.neurosolution.com/products/nsexcel/>.

Patients of the internal cohort were randomly divided into a training/cross-validation group (188 cases; 75%) and an internal validation group (63 cases; 25%; fig 1). The training/cross-validation group was used to train the network; in particular, 125 patients (50% of the entire study group) were randomly selected to train the network and the remaining 63 (25% of the entire study group) were used for internal cross validation.

In this study, a multilayered perceptron was constructed; a multilayered perceptron is a neural network that has at least three layers of neurones: one input layer, one output layer and at least one hidden layer. Neurones are tied together with weighted connections. Usually, the number of neurones in the input and output layers is determined by the amount of data available for learning: the number of attributes of the data (laboratory tests in this study) is equal to the number of input neurones and the number of outcomes is equal to the number of output neurones. The present multilayered perceptron had 10 input neurones: the input data selected for the development of the neural network were those available for both the internal and the external cohorts, and all were laboratory objective and reproducible variables—namely, aspartate aminotransferase (IU/l), total serum bilirubin (mg/dl), γ -glutamyl transpeptidase (IU/l), alkaline phosphatase (IU/l) serum creatinine (mg/dl), serum albumin value (g/dl), INR value, platelet count ($\times 10^3/\text{mm}^3$), white cell count ($\times 10^3/\text{mm}^3$) and haemoglobin concentration (g/dl; fig 2).

We found four neurones in the hidden layer and one output neurone (survival or death at 3 months). The learning rule used was back-propagation of error, which adjusts the internal parameters of the network over the repeated training cycles to reduce the overall error: once an input pattern is applied as a stimulus to the first layer, it is propagated through each upper layer until an output is generated; this output is compared with the desired output, and an error signal is calculated for each output unit. This error signal is then transmitted backwards across the net, and the weight of the connections in the network is updated to decrease the overall error of the net.¹¹

Training was terminated when the sum of square errors with respect to the cross-validation dataset of 63 patients was at a minimum. This cross validation was necessary because neural networks can be overtrained to recognise specific cases in a training set rather than learning general predictive characteristics; overtraining can lead to good performance on a training set but poor performance on independent testing sets. The activation function was used with continuous output on the

Table 1 Baseline characteristics of the internal and external cohort

Variable	Internal cohort (Bologna) n = 251	External cohort (London) n = 137	p Value
Demographic			
Age (years)	51.7 (8.5)	50.5 (12.5)	0.345
Sex (male)	206 (82%)	97 (71%)	0.015
Indication for LT			
Hepatitis B or C cirrhosis	196 (78.1%)	86 (62.8%)	0.001
Alcoholic liver disease	37 (14.7%)	19 (13.9%)	
Primary cholestatic disease*	18 (7.2%)	32 (23.4%)	
Laboratory values			
AST (IU/l)	97.2 (75.6)	89.4 (72.6)	0.369
Total bilirubin (mg/dl)	4.7 (3.4)	4.8 (3.3)	0.780
GGT (IU/l)	72.1 (68.4)	169.7 (119.4)	0.001
ALP (IU/l)	348.8 (207.0)	272.0 (232.8)	0.015
Creatinine (mg/dl)	0.94 (0.23)	1.36 (1.02)	0.001
Albumin (g/dl)	3.1 (0.5)	3.0 (0.7)	0.130
INR	1.65 (0.40)	1.28 (0.50)	0.001
Platelet count ($\times 10^3/\text{mm}^3$)	73.2 (51.8)	143.3 (128.5)	0.001
WBC ($\times 10^3/\text{mm}^3$)	4.5 (1.7)	5.2 (1.9)	0.276
Haemoglobin (g/dl)	11.8 (5.3)	11.1 (2.1)	0.093
MELD score	16.7 (4.8)	14.7 (6.9)	0.003

ALP, alkaline phosphatase; AST, aspartate aminotransferase; GGT, γ -glutamyl-transpeptidase; INR, international normalised ratio; LT, liver transplantation; MELD, model for end-stage liver disease; WBC, white cell.
 Continuous variables are reported as mean (SD).
 *Included primary biliary cirrhosis and primary sclerosing cholangitis.

interval (0–1), in which 0 = patient alive at 3 months and 1 = patient dead at 3 months so that ANN output ranged from 0 = probability of death 0% to 1 = probability of death 100%; transitional values were representative of different probabilities of death.

The ANN was trained 10 times. At the end of each training session, the network was tested and the prediction accuracy was calculated on the internal validation group of 63 patients who were not selected for training and whose outcome was unknown to the ANN. We then selected the best network in terms of accuracy and finally tested the ANN on the patients of

the external cohort (fig 1) to confirm its power to predict outcome and assess the possibility of applying the test to different subsets of patients.

The study protocol was in accordance with the Declaration of Helsinki and subsequent amendments.

STATISTICAL ANALYSIS

Continuous variables were expressed as mean (standard deviation (SD)) and compared using a two-tailed unpaired Student’s t test; categorical variables were compared using χ^2 analysis. Performance of the ANN prediction in the internal validation group and in the external cohort was tested using receiver-operating characteristic (ROC) curve analysis.¹² ROC analysis was also used to compare ANN and MELD performance using the Hanley and McNeil method. Finally, the ANN predictions for the internal validation group and the external cohort were taken together and expressed in terms of overall accuracy (sum of correct predictions divided by total predictions), sensitivity, specificity, positive predictive value and negative predictive value for several cut-off values considered.¹³ A value of $p < 0.05$ was considered significant in all the analyses.

Statistical analysis of continuous and categorical variables and ROC curve analysis were computed using MedCalc V.7.2.1.0 (MedCalc software, Mariakerke, Belgium).

RESULTS

Internal cohort

Table 1 shows the baseline characteristics of the internal cohort. The mean (SD) age was 51 (8.5) years, and the patients were predominantly men (82%). The most common indication for liver transplantation was cirrhosis of viral origin (196 patients; 78.1%) followed by alcohol-related chronic liver disease (37 patients; 14.7%) and cholestatic liver disease (18 patients; 7.2%). The mean (SD) MELD score of the internal cohort was 16.7 (4.8). Patients who survived had a lower total bilirubin, creatinine, INR value, white cell count and higher serum γ -glutamyl-transpeptidase, albumin value and haemoglobin concentration (table 2).

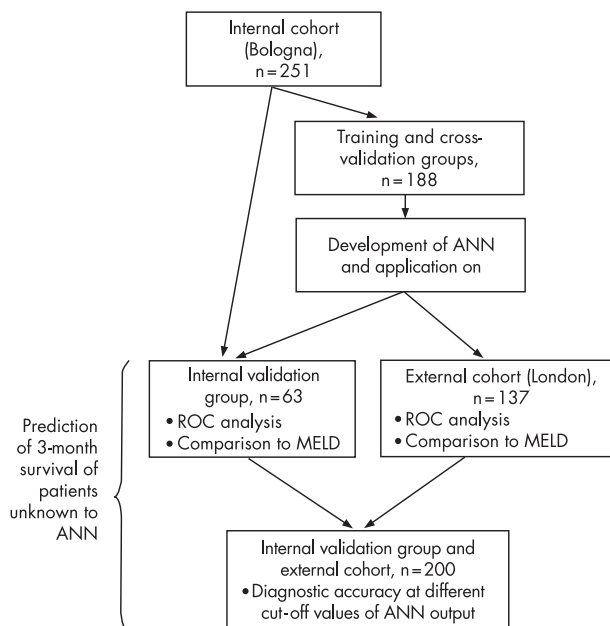


Figure 1 Datasets used in the development and validation of the artificial neural network (ANN). MELD, model for end-stage liver disease; ROC, receiver-operating characteristics.

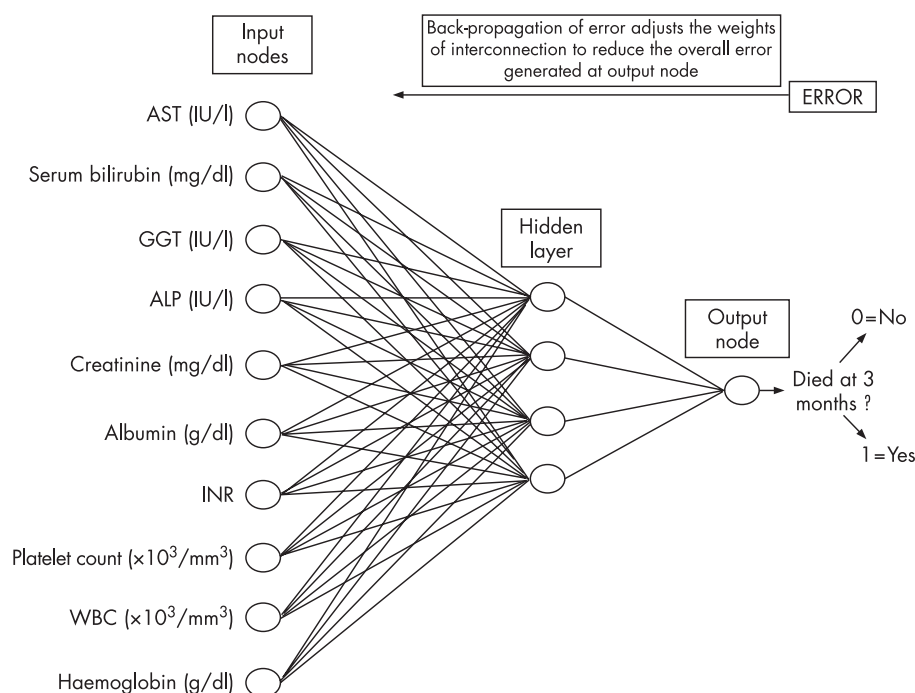


Figure 2 Schematic representation of the artificial neural network (ANN) model developed to predict 3-month mortality of patients with cirrhosis awaiting liver transplantation. ALP, alkaline phosphatase; AST, aspartate aminotransferase; GGT, γ -glutamyl-transpeptidase; INR, international normalised ratio; WBC, white cell count.

Of 188 patients, 27 (14.4%) in the training/cross-validation group and 11 of 63 (17.5%) in the internal validation group died within 3 months from the time they were listed for liver transplantation; this difference was not significant ($p = 0.696$). There was no difference in the median (SD) MELD score between the training/cross-validation group (16.8 (4.2)) and the internal validation group (16.7 (4.2); $p = 0.994$).

Performance of the ANN in predicting 3-month mortality in the training/cross-validation group was very high, with an area under the ROC curve (AUC) of 0.98 (95% confidence interval (CI) 0.94 to 0.99); in the same dataset the MELD had an AUC of 0.86 (95% CI 0.80 to 0.91) significantly lower than that of the ANN ($p = 0.002$). The performance of the ANN once applied to the internal validation group remained extremely high with an AUC of 0.95 (95% CI 0.86 to 0.99); the MELD AUC was 0.85 (95% CI 0.74 to 0.96), significantly lower than that of the ANN ($p = 0.032$; table 3).

External cohort

Table 1 shows the characteristics of the external cohort. The mean (SD) age was 50.5 (12.5) years and the patients were predominantly men (71%). The most common indication for liver transplantation was cirrhosis due to viral hepatitis (86

patients; 62.8%), followed by cholestatic liver disease (32 patients; 23.4%) and alcoholic liver disease (19 patients; 13.9%). The mean (SD) MELD score of the external cohort (14.7 (6.9)) was significantly lower than the score of the internal cohort (16.7 (4.8); $p = 0.003$).

We also observed significant differences in the percentage of men and patients with viral-related cirrhosis, as well as in the mean values of several laboratory tests between the internal and the external cohorts.

Similar to the internal cohort, patients who survived had a lower total bilirubin, creatinine, INR value and higher serum γ -glutamyl transferase and albumin value (table 2).

When the external cohort was submitted to the ANN, the AUC was 0.96 (95% CI 0.91 to 0.98); the performance of the MELD in predicting 3-month mortality in this cohort was significantly lower than that of the ANN ($p = 0.044$), with an AUC of 0.86 (95% CI 0.79 to 0.91; table 3).

Diagnostic accuracy of ANN of the internal validation group and external cohort

The internal validation group and external cohort consisted of 200 patients (172 alive and 28 dead at 3 months). Figure 3

Table 2 Variables of the internal cohort used to build the neural network.

Variable	Internal cohort (Bologna)			External cohort (London)		
	Survivors (n = 213)	Dead (n = 38)	p Value	Survivors (n = 120)	Dead (n = 17)	p Value
AST (IU/l)	93.7 (68.6)	116.6 (105.8)	0.207	84.2 (65.6)	116.1 (101.9)	0.253
Total bilirubin (mg/dl)	3.4 (2.9)	11.7 (9.6)	0.001	4.1 (5.4)	9.0 (8.5)	0.015
GGT (IU/l)	77.3 (63.9)	43.1 (25.4)	0.001	186.7 (43.3)	81.0 (79.5)	0.007
ALP (IU/l)	357.1 (219.0)	302.9 (110.1)	0.137	281.1 (152.5)	224.7 (204.9)	0.469
Creatinine (mg/dl)	0.92 (0.21)	1.02 (0.29)	0.011	1.28 (1.23)	1.79 (1.64)	0.044
Albumin (g/dl)	3.1 (0.4)	2.9 (0.4)	0.016	3.1 (0.6)	2.5 (0.8)	0.001
INR	1.57 (0.31)	2.11 (0.54)	0.001	1.21 (0.37)	1.68 (0.83)	0.017
Platelet count ($\times 10^3/\text{mm}^3$)	74.7 (53.8)	64.6 (37.8)	0.267	149.1 (135.3)	112.7 (103.5)	0.294
WBC ($\times 10^3/\text{mm}^3$)	4.3 (1.6)	5.5 (2.2)	0.002	6.1 (5.1)	7.5 (6.3)	0.236
Haemoglobin (g/dl)	12.1 (5.7)	10.2 (1.3)	0.038	11.1 (1.9)	10.6 (2.4)	0.229

ALP, alkaline phosphatase; AST, aspartate aminotransferase; GGT, γ -glutamyl-transpeptidase; INR, international normalised ratio; WBC, white cell. Continuous variables are reported as mean (SD).

Table 3 Receiver-operating characteristic analysis of the artificial neural network in comparison with the model of the end-stage liver disease performance

	ANN		MELD score		p Value*
	AUC	95% CI	AUC	95% CI	
Training/cross validation	0.98	0.94 to 0.99	0.86	0.80 to 0.91	0.002
Internal validation	0.95	0.86 to 0.99	0.85	0.74 to 0.96	0.032
External cohort	0.96	0.91 to 0.98	0.86	0.79 to 0.91	0.044

ANN, artificial neural networks; AUC, area under the curve; MELD, model for end-stage liver disease.
*p values of each ANN subgroup compared with the MELD score of the same subgroup (Hanley–McNeil method).

shows the sensitivity, specificity, positive predictive value, negative predictive value and overall accuracy of the ANN output in the internal validation group and external cohort at different cut-off values. Using the lower cut-off values of 0.03, the model had a sensitivity of 100%, and a negative predictive value of 100%: lower values of ANN output indicate a 100% probability of patients being alive at 3 months. Above this threshold, the probability of death starts to increase from 41.2% (positive predictive value) up to 90.9% when the ANN output reaches the value of 1. The overall accuracy remains high for all the cut-offs considered, ranging from 82% to 91%.

DISCUSSION

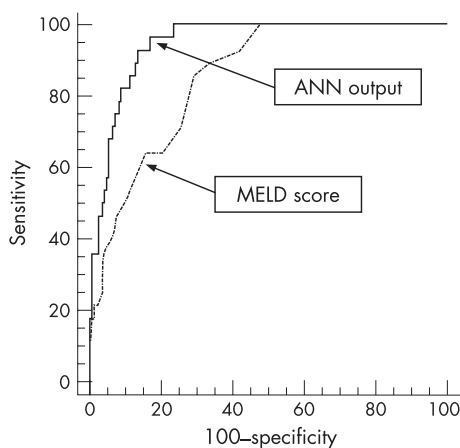
The MELD represents the milestone of the present policy of allocation of donor organs that prioritises the sickest patients on the waiting list for liver transplantation. Being based on simple and objective biochemical parameters, the MELD avoids the possibility of manipulating the process of organ allocation that was consistent when it was based on waiting time, location of the patient (hospitalised *v* at home) and subjective clinical parameters. However, the MELD score will not accurately

predict short-term mortality in 15%–18% of the patients with chronic liver disease listed for liver transplantation; in practical terms, many patients undergo transplantation too early whereas others die awaiting a donor. The search for more precise systems to predict the prognosis of liver transplant candidates is therefore justified.¹⁴

In this study, the ANN was superior to the MELD scoring system in predicting the 3-month mortality of patients with end-stage liver disease listed for liver transplantation. The MELD score was developed by Malinchoc using a proportional hazard model to obtain the relative risk related to the modification of a defined variable and then applied to a logistic regression to predict the probability of death within 3 months of patients with cirrhosis undergoing transjugular intrahepatic portosystemic stent shunt.² Recent studies have shown that neural network analysis is potentially more successful than traditional statistical techniques when the importance of a given prognostic variable is expressed as a complex unknown function of the value of the variable, or when the prognostic effect of a variable is influenced by other prognostic variables in a complex multidimensional non-linear function.^{15–19} This condition is present in a complex biological system such as end-stage liver disease. On the basis of these suggestions, we constructed an ANN using the clinical parameters that are most commonly recorded when a patient is listed for liver transplantation, and we tested its ability to predict the 3-month mortality.

In his original study, Malinchoc showed that serum bilirubin, serum creatinine, INR and cause of cirrhosis are independent risk factors of mortality of patients with cirrhosis undergoing the transjugular intrahepatic portosystemic stent shunt procedure; the effect of adding other important clinical features to the score, such as the presence of refractory ascites, variceal bleeding, spontaneous bacterial peritonitis and encephalopathy, was tested, and produced no significant improvement in the prediction of 3-month mortality.^{2 3 5} More recently, it was shown that the accuracy of the MELD in the liver transplantation setting is improved by the addition of serum sodium values to the formula; however, the increase of the c statistic obtained by adding sodium values was only slightly >1% (MELD score c statistic 0.894 *v* MELD score plus hyponatraemia 0.905).²⁰

The simple addition of further prognostic variables does not seem to substantially improve the fit of a model that is developed with conventional discriminant analysis. In the context of a biological system, such as end-stage liver disease, a neural network can process more information, providing a different analysis of prognostic factors than traditional statistical techniques. In our report, the c statistic achieved by the ANN when internal and external cohorts were considered was some 10% greater than that obtained with the MELD score. Notably, in our series the performance of MELD was in accordance with what was reported in other series, with an accuracy as high as 0.86; this observation, while confirming the efficacy of the MELD system as a predictor of short-term



ANN output cut-off points	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)	Accuracy (%)
0.03	100	76.7	41.2	100	82.0
0.1	96.4	83.1	48.2	99.3	85.0
0.3	85.7	87.2	52.2	97.4	87.0
0.6	82.1	91.3	60.5	96.9	89.5
0.9	67.9	94.8	67.9	94.8	91.0
1.0	35.7	99.4	90.9	90.5	90.0

PPV: positive predictive value
NPV: negative predictive value

Figure 3 Diagnostic accuracy at different cut-off values of the artificial neural networks (ANN) output in the internal validation group and external cohort. MELD, model for end-stage liver disease.

mortality, adds further value to the performance of the ANN which was better than that of an already highly reliable system.

As the network used in this study was constructed using 10 variables, each group should have 100 patients to avoid the risk of overfitting of the data. This was not fully achieved for the internal validation group, which consisted of 63 patients. However, the model was confirmed in the external validation group of >100 people.

In this study, the ANN was shown to have the best accuracy reported in the literature in predicting the mortality of patients with end-stage liver disease; in a liver transplantation setting the ANN had an outstanding ability to identify patients who will die of liver failure within 3 months among those who are awaiting transplantation. The ANN can provide a score ranging from 0 (no probability of death) to 1 (100% probability of death) to assess the risk of death on the waiting list for liver transplantation. In clinical practice, the software used for the present neural network allows the creation of a new mask that can be incorporated in a website easily accessible to everyone (as in the UNOS website, where a mask was created for MELD calculation). The possibility of identifying those patients who are to undergo transplantation more urgently by using the ANN could optimise organ allocation, and is therefore of paramount importance.

When the ANN was tested on the external cohort, its accuracy remained similar to that observed in the internal validation group and higher than that of the MELD score. As some of the characteristics of the patients in the external cohort, such as sex distribution, aetiology of liver disease and laboratory tests values, differed from the internal validation group, we conclude that the ANN can offer reproducible results.

Although the addition of clinical variables, such as spontaneous bacterial peritonitis, degree of encephalopathy or ascites, might have improved the prognostic capability of the network analysis, we decided not to include them because their assessment is subjective and could therefore be arbitrary, particularly in a centralised or shared organ-allocation system.

In conclusion, we provide some evidence that ANN are valid and potentially reliable tools to determine the short-term prognosis of patients with end-stage liver disease. The ANN that we developed was superior to the MELD scoring system in predicting the 3-month survival of patients listed for liver transplantation this was also the case when it was applied to patients referred to different transplant centres. The ANN could be used as a disease severity index to better prioritise patients for organ allocation, reducing the number of deaths on the waiting list for liver transplantation.

Authors' affiliations

A Cucchetti, M Vivarelli, G La Barba, A D Pinna, Dipartimento di Discipline Chirurgiche, Rianimatorie e dei Trapianti, University of Bologna, S Orsola-Malpighi Hospital, Bologna, Italy

F Piscaglia, L Bolondi, Dipartimento di Medicina Interna e Gastroenterologia, University of Bologna, S Orsola-Malpighi Hospital, Bologna, Italy

N D Heaton, S Phillips, M Rela, Department of Liver Transplantation and Hepatobiliary Surgery, King's College Hospital, Denmark Hill, London, UK

M R Foxton, J O'Grady, Institute of Liver Studies, King's College Hospital, Denmark Hill, London, UK

Competing interests: None.

REFERENCES

- 1 **Anonymous**. Organ procurement and transplantation network-HRSA. Final rule with comment period. *Federal Register* 1998;**63**:16296-338.
- 2 **Malinchoc M**, Kamath PS, Gordon FD, et al. A model to predict poor survival in patients undergoing transjugular intrahepatic portosystemic shunts. *Hepatology* 2000;**31**:864-71.
- 3 **Kamath PS**, Wiesner RH, Malinchoc M, et al. A model to predict survival in patients with end-stage liver disease. *Hepatology* 2001;**33**:464-70.
- 4 **Wiesner RH**, Edwards E, Freeman R, et al. United Network for Organ Sharing Liver Disease Severity Score Committee. Model for end-stage liver disease (MELD) and allocation of donor livers. *Gastroenterology* 2003;**124**:91-6.
- 5 **Wiesner RH**, Mc Diarmid SV, Kamath PS, et al. MELD and PELD: application of survival models to liver allocation. *Liver Transpl* 2001;**7**:567-80.
- 6 **Cross S**, Harrison RF, Kennedy RL. Introduction to neural networks. *Lancet* 1995;**346**:1075-9.
- 7 **Baxt WG**. Application of artificial neural network to clinical medicine. *Lancet* 1995;**346**:1135-8.
- 8 **Forsstrom JJ**, Dalton KJ. Artificial neural networks for decision support in clinical medicine. *Ann Med* 1995;**27**:509-17.
- 9 **Azulay D**, Linhares MM, Huguet E, et al. Decision for retransplantation of the liver: an experience- and cost-based analysis. *Ann Surg* 2002;**236**:713-21.
- 10 **United Network for Organ Sharing (UNOS)**. Policy 1999 Available at <http://www.unos.org> (accessed 27 Oct 2006).
- 11 **Rumelhart DE**, Hinton GE, Williams RL. Learning representation by back-propagating errors. *Nature* 1986;**323**:533-6.
- 12 **Hanley JA**. Receiver operating characteristic (ROC) methodology: the state of the art. *Crit Rev Diagn Imaging* 1989;**29**:307-35.
- 13 **Altman DG**, Bland MJ. Diagnostic tests 2: predictive values. *BMJ* 1994;**309**:102.
- 14 **Wiesner RH**. Evidence-based evolution of the MELD/PELD liver allocation policy. *Liver Transpl* 2005;**11**:261-3.
- 15 **De Laurentiis M**, Ravdin PM. A technique for using neural network analysis to perform survival analysis of censored data. *Cancer Lett* 1994;**77**:127-38.
- 16 **De Laurentiis M**, Ravdin PM. Survival analysis of censored data: neural network analysis detection of complex interaction between variables. *Breast Cancer Res Treat* 1994;**32**:113-18.
- 17 **Rumelhart D**, McClelland J, eds. *Parallel distributed processing: explorations in the microstructures of cognition*. Cambridge, MA: MIT Press, 1986.
- 18 **Lapuerta P**, Rajan S, Bonacini M. Neural networks as predictors of outcomes in alcoholic patients with severe liver disease. *Hepatology* 1997;**25**:302-7.
- 19 **Banerjee R**, Das A, Ghoshal UC, et al. Predicting mortality in patients with cirrhosis of liver with application of neural network technology. *J Gastroenterol Hepatol* 2003;**18**:1054-60.
- 20 **Ruf AE**, Kramers EK, Chavez LL, et al. Addition of serum sodium into the MELD score predicts waiting list mortality better than MELD alone. *Liver Transpl* 2005;**11**:336-43.