ANNALS OF SURGERY
Vol. 219, No. 4, 408–415

Predicting Outcomes After Liver Transplantation

A Connectionist Approach

Howard R. Doyle, M.D., Igor Dvorchik, M.Sc., Sandi Mitchell, R.Ph., M.S.I.S., Ignazio R. Marino, M.D., Fredrick H. Ebert, B.Sc., John McMichael, B.Sc., and John J. Fung, M.D., Ph.D.

From the Section of Computational Medicine, the Transplantation Institute, University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania

Objective
The authors sought to train an artificial neural network to predict early outcomes after orthotopic liver transplantation.

Summary Background Data
Reliable prediction of outcomes early after liver transplantation would help improve organ use and could have an impact on patient survival, but remains an elusive goal. Traditional multivariate models have failed to attain the sensitivity and specificity required for practical clinical use. Alternate approaches that can help us model clinical phenomena must be explored.

One such approach is the use of artificial neural networks, or connectionist models. These are computation systems that process information in parallel, using large numbers of simple units, and excel in tasks involving pattern recognition. They are capable of adaptive learning and self-organization, and exhibit a high degree of fault tolerance.

Methods
Ten feed-forward, back-propagation neural networks were trained to predict graft outcomes, using data from 155 adult liver transplants. The data included information that was available by the second postoperative day. Ten separate training and testing data subsets were prepared, using random sampling, and the ability of the different networks to predict outcomes successfully was evaluated using receiver operating characteristic (ROC) curve analysis.

Results
Four of the networks showed perfect discrimination, with an area under the ROC curve (A_z) of 1.0. Two other networks also had excellent performance, with an A_z of 0.95. The sensitivity and specificity of the combined networks was 60% and 100%, respectively, when using an output neuron activation of 0.6 as the cutoff point to decide class membership. Lowering the cutoff point to 0.14 increased the sensitivity to 77%, and lowered the specificity to 96%.

Conclusions
These results are encouraging, especially when compared to the performance of more traditional multivariate models on the same data set. The robustness of neural networks, when confronted with noisy data generated by nonlinear processes, and their freedom from a priori assumptions regarding the data, make them promising tools with which to develop predictive clinical models.

Outcome prediction is becoming increasingly more important in medicine, for reasons that range from the purely academical to the overtly fiscal. Predictive models must perform well when applied to individual cases, or clinicians will be reluctant to use them, and any theoretical benefit in terms of improved patient care and resource use will go unrealized.

Liver transplantation is one field that would benefit greatly from such models. We must be able to determine early whether a transplanted organ is destined to fail, because early intervention can ameliorate the high morbidity and mortality rates that accompany retransplantation. Also, the large discrepancy between the increasing demand for organs and their stagnant supply,[1] coupled with the high cost of these procedures,[2] make it imperative that we optimize the use of such scarce resources.

Early prediction of outcomes after liver transplantation still is not feasible, except in general terms. The reasons for this are complex, and may involve a combination of our incomplete understanding of all the processes involved, relatively small patient samples, and the use of modeling techniques poorly suited for the task.[3]

Artificial neural networks are computation systems, implemented either in hardware or software, that mimic the computation abilities of biologic systems by using large numbers of simple, interconnected, artificial neurons. These neurons take information from sensors or other artificial neurons, perform simple operations on the data, and pass the results on to other artificial neurons. They exhibit adaptive learning, self-organization, and fault-tolerance. They also can operate in real time, and are inserted easily into existing technology.[4]

Neural networks have been used to solve problems in a wide range of fields, such as handwritten character recognition,[5] sonar signal processing,[6] image reconstruction,[7] robotics,[8] and nucleic acid sequence prediction.[9] Although their development was associated closely with the neurosciences and they have been used extensively to model the nervous system at different levels of complexity,[10,11] neural networks have not made much of an impact in clinical medicine. However, there is some evidence that the well-known strengths of neural computation can be applied successfully in a clinical setting.[12-16] The purpose of this study was to train artificial neural networks to predict patient and graft outcomes after liver transplantation, using only data available in the early post-transplant period, and to evaluate their predictive performance. Our results suggest that neural networks are well suited for this task, and may perform better than other, more traditional, approaches.

Supported by Project Grant No. DK 29961 from the National Institutes of Health, Bethesda, Maryland.

Address reprint requests to Howard R. Doyle, M.D., 3601 Fifth Avenue, Suite 5C, Pittsburgh, PA 15213.

Accepted for publication October 11, 1993.

CASE MATERIAL AND METHODS

Patient Population

From January to August 1992, 155 adult patients underwent 162 liver transplants at Pittsburgh's Presbyterian University Hospital; 149 patients survived more than 24 hours. These patients were entered in a prospective study aimed at analyzing factors that could be used for early prediction of outcomes after liver transplantation, and their progress was followed for 3 months. A detailed description of that study has been reported.[3] One patient died of intractable supraventricular arrhythmias, 23 days after surgery, that resulted from trauma to the atrial conduction system after a direct anastomosis between the donor suprahepatic vena cava and the recipient's heart. At the time of death from cardiac arrest, the graft function was normal. This patient was excluded from the analysis, leaving 148 patients undergoing 155 separate transplant events, which form the basis for this report.

Because these data were obtained during an observation study, in which patients were treated according to our established clinical protocols, Institutional Review Board approval was not necessary.

DEFINITIONS

Graft Failure

Graft failure is defined as patient death or retransplantation, within 3 months of surgery, in patients that survived at least 24 hours after the operation.

Primary Non-Function

Primary non-function, lacking a technical complication, is a graft that never demonstrates evidence of initial function, so that retransplantation must be carried out within 2 weeks of the original operation, or the patient succumbs to liver failure before a suitable graft can be obtained.

Sepsis

Sepsis is systemic response to infection, manifested by two or more of the following:

Temperature > 38 C or < 36 C;
Heart rate > 90 beats/min;
Respiratory rate > 20 breaths/min or $PaCO_2$ < 32 torr;

White blood cell count > 12,000 cell/mm^3, < 4000 cells/mm^3, or > 10% immature (band) forms.[17]

Severe Sepsis

Severe sepsis is sepsis associated with organ dysfunction, hypoperfusion, or hypotension.[17]

NETWORK INPUT PARAMETERS

Input parameters were chosen based on results obtained in a previous study,[3] our changing practice regarding preoperative and postoperative clinical monitoring, and the clinical judgment of the investigators. The input patterns were formed by 19 parameters (Table 1). These parameters can be subdivided in two groups, according to whether the information was obtained on the day of the transplant or the first 2 postoperative days.

The parameters from the day of the transplant included the following: patient's age, serum bilirubin level, serum creatinine level, prothrombin time, need for pretransplant mechanical ventilation, whether the index transplant was a retransplantation, and the peak intraoperative serum lactate level. The same laboratory measurements and the serum aspartate aminotransferase (AST) were obtained postoperatively as part of the patient's routine biochemical monitoring. The need for fresh frozen plasma during the preceding 24 hours also was recorded to help interpret changes in the prothrombin time.

Table 1. PARAMETERS USED TO TRAIN THE NEURAL NETWORKS

Age
Retransplantation
Need for preoperative mechanical ventilation
Preoperative total serum bilirubin
Preoperative serum creatinine
Preoperative prothrombin time
Peak intraoperative serum lactate
AST POD 1
Total serum bilirubin POD 1*
Serum creatinine POD 1*
Serum lactate POD 1*
Prothrombin time POD 1*
FFP 1
AST POD 2*
Total serum bilirubin POD 2*
Serum creatinine POD 2*
Serum lactate POD 2*
Prothrombin time POD 2*
FFP 2

AST = aspartate aminotransferase; POD = postoperative day; FFP = whether the patient received fresh frozen plasma in the previous 24 hours.

* Relative change, calculated as $\frac{V_c - V_p}{V_p}$, where V_c = current value and V_p = value on previous day.

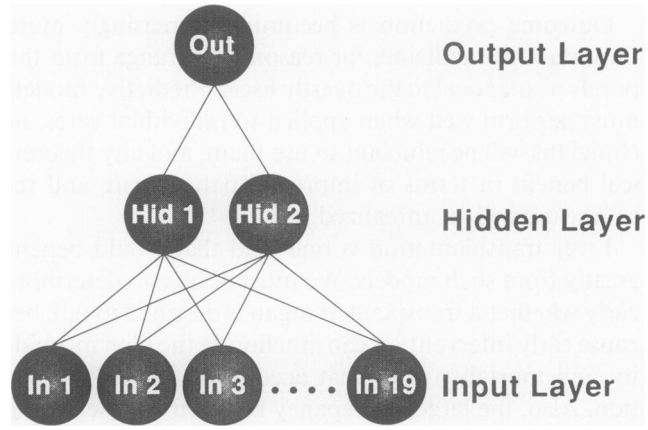

Figure 1. Architecture of the neural networks used in the present study. The network consists of three layers, with 19 neurons in the input layer, 2 neurons in the hidden layer, and 1 neuron in the output layer.

Continuous variables were presented to the network as both the initial value and measures representative of the magnitude of change in the early postoperative period. These consisted of the relative change in a given value with respect to that of the previous day (Table 1). All the available inputs were then scaled linearly between 0.1 and 1.0. Binary variables were given values of 0.1 or 1.0. If an observation was missing, whether binary or continuous, it was assigned the value of 0.

NETWORK ARCHITECTURE

We used a feed-forward, back-propagation network[18] with three layers. The input layer had 19 neurons; there were 2 neurons in the hidden layer, and one neuron in the output layer (Fig 1). The concept of a neuron, or node, is a higher-level abstraction that encompasses both certain values and a set of operations that are performed on those values (Fig 2). The neuron receives an input signal, or a set of input signals, coming from other neurons or input devices. After adding the signals received over one operating cycle, plus or minus a threshold value, the result is passed through an activation function, also known as a transfer function. The result of this is the activation of the neuron, which is dependent functionally on the earlier summed inputs. This activation, or output, is passed on by the neuron to all other neurons to which it is connected, after multiplying it by values that represent the connection weights.[19] Although there are a number of functions that are suitable candidates, the most popular transfer functions are the sigmoids, which are continuous, real-valued functions whose do-

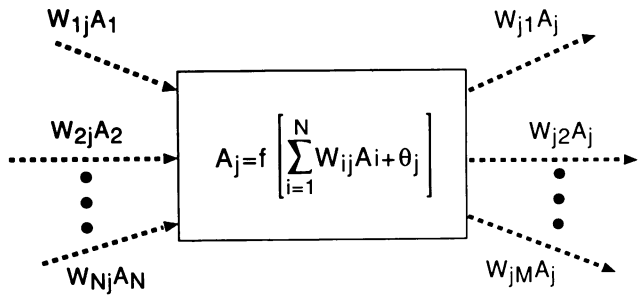

Figure 2. Schematic representation of a single neuron. The neuron receives inputs from other neurons or input devices. The inputs (A_1–A_N) are multiplied by the corresponding connection weights (W_{1j}–W_{Nj}) and then totalled. A constant term (θj), or bias, is then added, and the result is passed through the transfer function. The output of the transfer function is the output of the neuron. Adapted with permission from Maren AJ.[19]

main is the real numbers; their derivative is always positive, and their range is bounded. The most commonly used sigmoid function is the logistic,

$$f(x) = \frac{1}{1 + e^{-x}},$$

which is the one used in this study. As can be seen from Figure 1, the neurons are connected fully with those of the preceding and subsequent layers, but there are no lateral connections in this particular architecture.

LEARNING ALGORITHM

Feed-forward, back-propagation networks can learn to perform a mapping between input and output patterns. Their learning algorithm was described initially by Paul Werbos,[20] but it only became popular after the work of Rumelhart et al.[18] Learning is supervised, and training usually begins by setting the connection weights randomly, although initial weights can be determined by other methods. Then the network is presented with sets of pairs of input and output patterns. The input pattern, or vector, is used by the network to produce its own output vector, which is then compared with the target vector. If they are different, it uses the back propagation of errors, or generalized delta rule,[21] to adjust the connection weights to minimize the error. This readjustment can be performed after each training pattern is presented ("on-line" or continuous updating), or after a complete set of training patterns have been processed, i.e., after a completed epoch ("batch" or periodic updating), and continues until the weights settle into a stable state. When a network reaches a stable state, it is said to have converged; however, convergence does not guarantee that useful learning has taken place.

In its most basic form, back propagation is a gradient descent algorithm. It can be slow for some applications, and it scales up poorly as tasks become larger and more complex.[22] In this study, the algorithm used was quickprop, a variation of standard back propagation that has been found to be much faster, and that appears to scale much better.[22] The implementation of the quickprop algorithm used in this study (NevProp) was developed at the Center for Biomedical Modeling Research, University of Nevada, Reno.

TRAINING STRATEGY

After preliminary experiments, a combination of network architecture and learning parameters that consistently produced satisfactory results was chosen. The 155 patterns were divided into two groups, a training set, composed of 89% of the total patterns, and a testing set, composed of the remaining 11% of the patterns. These sets were made by random selection from each subgroup (failed and successful grafts). Enough cases were selected so that their proportions would be maintained approximately in both the training and testing sets. The order of presentation of the training patterns, with respect to their outcome, was also random. The network was trained on the first set, and its performance was tested on the second set. Training was continued until the best generalizing weights were obtained.

Because an individual result could be skewed by a particularly favorable or unfavorable training/testing set combination, ten different sets were made. The patterns were selected at random. Then these were used to train and test ten different networks, and their performance was analyzed separately. All randomizations were performed with the use of a random number generator.

ANALYSIS OF NETWORK PERFORMANCE

Receiver operating characteristic (ROC) curve analysis[23] was used to evaluate the performance of the trained neural networks. The predictive performance was assessed in terms of the true-positive and false-positive fractions. The true-positive fraction is the fraction of actual failed grafts that were predicted correctly as failures (i.e., the sensitivity). The false-positive fraction is the fraction of actual successful grafts that were predicted incorrectly to fail (1 − specificity). The area under the ROC curve (A_z) was used as an index of performance. Calculations were done with software (Labroc1) provided by Dr. Charles Metz from the Department of Radiology, University of Chicago.

RESULTS

There were 2945 data points (19 variables and 155 cases), of which 18 (0.6%) were missing values, and were

Table 2. INDIVIDUAL PERFORMANCE OF THE TRAINED NEURAL NETWORKS, AS MEASURED BY THE AREA UNDER THE ROC CURVE (A_z)

Network	A_z
1	1.0
2	0.89
3	0.86
4	0.87
5	1.0
6	0.95
7	0.88
8	0.95
9	1.0
10	1.0

An area of 1.0 indicates perfect discrimination.

assigned a value of zero at the time of normalization. Of 155 grafts analyzed, 135 (87.1%) were successful according to our definition, and 20 (12.9%) failed. Eleven failures were attributed to patient death, and the remaining nine needed retransplantation. The specific causes of failure were as follows:

Primary Non-Function. There were three cases of primary non-function.

Severe Sepsis. Nine grafts failed because of patient death from severe sepsis 2 to 64 days after transplantation.

Pancreatitis. One patient died of acute necrotizing pancreatitis 30 days after transplantation.

Arrhythmias. One patient died after developing refractory ventricular fibrillation, 3 days after surgery. He had no predisposing risk factors.

Ischemic Injury. Five grafts were lost to severe ischemic injury 16 to 90 days after transplantation.

Rejection. One graft was lost to uncontrolled acute rejection, with severe central venulitis, 33 days after transplantation.

Network Performance. Table 2 shows the individual performance of the ten neural networks, as measured by the area under the ROC curve (A_z). This is a good indication of overall performance, with an area of 1.0 indicating that a test has perfect discrimination, and an area of 0.5 indicating that it performs no better than chance.[23] Four of the networks showed perfect discrimination on the testing subset, and two others had an A_z of 0.95. The worst performance was that of network #3, with an A_z of 0.86.

To get an idea of the overall performance of the ten neural networks, we combined all the predictions into one set, and performed ROC curve analysis on it. Figure 3 shows the resulting ROC curve for this combined set. The area under the curve is 0.90. But more importantly, the false-positive fraction remains at zero until the true-positive fraction reaches 0.6. This can be better appreciated in Table 3, in which the false-positive and true-positive fractions change as we vary the cut-off point that we use to determine whether a graft will fail. If we set such cut-off point at an output neuron activation of 0.6, our false-positive fraction is zero and our true positive fraction is 0.6. In other words, the specificity is 100% and the sensitivity is 60%. Setting the cut-off point at a neuron activation of 0.14 makes the false-positive fraction go up to 0.04, while the true-positive fraction increases to 0.77 (96% specificity and 77% sensitivity).

DISCUSSION

Although relatively new to clinical medicine, neural computation has evolved slowly since 1943, when Warren McCulloch and Walter Pitts published a seminal paper that profoundly influenced later computer scientists.[24] In it, they attempted to describe the inner workings of the nervous system by using primitive computing elements that were mathematical abstractions of actual neurons and their connections, as they understood them at the time. The so-called "McCulloch-Pitts neuron" is a binary device (it can exist in one of two states) with a fixed threshold; it receives excitatory or inhibitory inputs and carries out its computations in discrete periods of

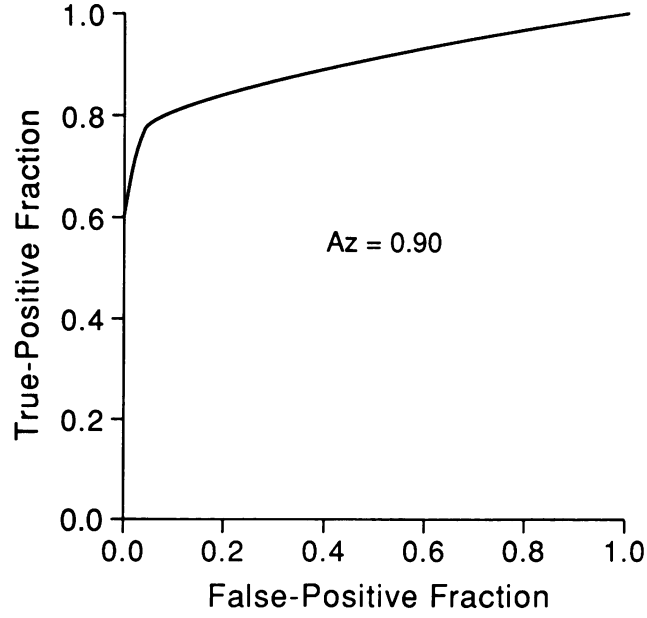

Figure 3. Receiver operating characteristic (ROC) curve of the combined neural networks.

Table 3. COMBINED PERFORMANCE OF THE 10 NEURAL NETWORKS AT VARIOUS CRITICAL OUTPUT NEURON ACTIVATIONS

Cut-Off Value	FPF	TPF
0.99	0.0	0.20
0.79	0.0	0.40
0.60	0.0	0.60
0.14	0.04	0.77
0.02	0.15	0.84
0.01	0.33	0.88

FPF = false positive fraction; TPF = true positive fraction.

time. If the sum of its excitatory inputs exceeds the threshold, the neuron becomes activated (i.e., "fires"). If the neuron receives at least one inhibitory input, it will remain inactive during that time cycle. Any finite logical function can be duplicated by a given network composed of these all-or-none neurons.

As important as their model was, the McCulloch-Pitts neuron oversimplified neural electrical activity. More importantly, the strength of the inputs did not change as a function of experience; therefore, the network could not learn. This latter deficiency was overcome by the advent of the perceptron,[25] the first real artificial neural network, which introduced several concepts currently in use. The perceptron generated much enthusiasm because it was a learning machine that could be implemented in the hardware available at the time, and it displayed adaptive behavior. It also could respond correctly to input patterns it had not seen before, i.e., it could generalize. But it had serious limitations[26] that led eventually to disenchantment with this line of research. Work in this field came practically to a halt until public attention was drawn to it again by Hopfield in 1982.[27]

The connectionist approach has been used with success in medical imaging and signal processing.[28] Medical imaging is an area that can naturally take advantage of the pattern recognition capabilities of neural networks. This includes interpretation of mammograms[16] or chest x-rays[29] based on features extracted by the reader, and automatic tumor classification from analysis of ultrasound images of the eye.[30] Signal processing applications range from electrocardiogram trace classification[31,32] to electroencephalogram analysis to determine depth of sedation.[14] The adeptness of neural networks at classification tasks also has been exploited by some researchers to try to develop clinical diagnostic systems as an alternative to more traditional expert systems. Mulsant used a four-layer back-propagation network trained to diagnose dementia, and achieved a 77% agreement with the diagnosis made by the clinicians.[33] Furlong et al.,[13] trained a network to classify patients in two groups, acute myocardial infarction (AMI) and no infarction (no-AMI), based on analysis of serial enzyme determinations. The network made the correct diagnosis, as validated by autopsy findings, in 92% of cases of AMI, and 67% of cases of no-AMI. Baxt also used a neural network to diagnose AMI in an emergency department, using historical data, physical signs, and electrocardiographic findings as inputs.[12] In a prospective comparison, physicians had a diagnostic sensitivity of 77.7% and a specificity of 84.7%, whereas the trained neural network had a sensitivity of 97.2% and specificity of 96.2%.

Outcome prediction can be viewed as another classification task, with individual patterns determining class membership, and it is logical to try to use neural networks for this purpose. This approach already has been found to be at least as good as Cox regression in predicting breast cancer relapse,[15] and may be useful in identifying patients who will not survive to discharge after cardiopulmonary resuscitation.[34] Our results indicate that neural networks can be trained successfully to predict patient and graft outcomes following liver transplantation. These predictions can be made as early as 48 hours after the operation, and the system takes easily obtained preoperative data and routine follow-up biochemical information as inputs (Table 1). We trained ten different networks of the same architecture, using different training and testing data sets selected at random from the same patient population and evaluated their performance using ROC curve analysis. As shown in Table 2, in four of those networks, the area under the ROC curve was 1.0 — i.e., perfect discrimination. The combined performance of the ten networks also was good, with an area under the ROC curve of 0.9 (Fig. 3). Table 3 shows the true-positive and false-positive fractions of the combined networks while we vary the output neuron activation value that is used as a cutoff point to decide class membership. Using an activation of 0.6 as a cutoff, the true-positive fraction is 0.6 (60% sensitivity), and the false-positive fraction is 0.0 (100% specificity). If the cutoff point is lowered to 0.14, the true-positive fraction increases to 0.77 (77% sensitivity), and the false-positive fraction increases to 0.04 (96% specificity). This high degree of specificity would make this model attractive in clinical practice, where difficult and irreversible decisions often are made in anticipation of poor outcomes. These neuron activation values are not probabilities, although they are often (loosely) interpreted as such.

Although preliminary, these results are encouraging. Recently, we completed an analysis of early predictive factors after liver transplantation, using the same patient population that we report in this study.[3] We found that a number of parameters correlated with early outcomes, but none were discriminating enough to make accurate

predictions. Even when several parameters were entered into stepwise logistic regression models, their predictive performance was relatively poor, the best model having a sensitivity of 75%, but a specificity of only 87%. Direct comparisons between the two methods are not straightforward, especially because neural networks allow us to use data in ways that often violate the underlying assumptions of traditional statistical techniques. Also, neural networks are much more robust when it comes to handling noise, and allow us a certain degree of freedom when confronted with missing data or simple measurement errors. The incidence of missing values in this series was small, 0.6%, but this led to the exclusion of some cases from the logistic regression model; the neural networks incorporated these cases after substituting zero for the missing value. There are other strategies available when dealing with missing values, such as substituting the class mean or median, but because of the small number of missing data points, we did not attempt a comparison between different methods. We plan to do such a comparison in the future, together with the deliberate introduction of noise, which can help train networks that are less brittle under actual use. Considering the different nature of the two methods, it is clear that neural networks compare favorably with logistic regression, especially because the performance of the networks was measured against test data sets (i.e., data not used during training) and the performance of the logistic regression model was estimated on the same data that was used to generate the model.[3]

Standard multilayered feed-forward networks can approximate virtually any function of interest to any desired degree of accuracy, provided enough hidden units are available and the relation between input and target patterns is deterministic, rather than stochastic.[35] There is a temptation, when setting up the network's architecture, to add a large number of hidden neurons. This increases the "processing power" of the network, and makes it more likely that the network will learn the training set. However, it also increases the probability that the network will soon learn all the relevant facts and start learning the noise that is present in the data, which will lead to poor generalization. We can overfit a neural network, just as we can overfit a regression model. In light of that, it also is encouraging that the architecture of the network used in this study involved only two hidden neurons. This is a good indication that the network is using a few robust features to perform its classification, rather than memorizing the idiosyncrasies present in the data.

As powerful as connectionist models are, we must be cautious before embracing them, and we must be aware of their pitfalls. With relatively small data sets, leaving out even a small number of patterns for subsequent testing can have important consequences for network performance. This is illustrated in Table 2, which shows that with different training/testing data set combinations, we obtain networks that perform very differently — some discriminate test cases perfectly while others do not perform as well. We also can introduce a bias during training when we continue to train until the network performs well on the testing set, in essence fitting the model to the testing set. We will need larger training and testing sets, and a separate cross-validation group, to confirm our initial findings. We also will need a larger group so that we can discriminate between patient death and simple organ failure. But even with these caveats, we believe that neural networks constitute a group of promising tools with which to build clinical prediction models. Their freedom from a priori assumptions regarding the data, their robustness in the face of noise, and their ability to solve nonlinear separable problems, make them well-suited for the clinical arena.

References

1. Marino IR, Doyle HR, Kang YG, et al. Multiple organ procurement. *In* Ayres SM, Grenvik A, Hollbrook PR, Shoemaker WC, eds. Textbook of Critical Care. 3rd ed. Philadelphia: WB Saunders (in press).
2. Evans RW, Mannienen DL, Dong FB. An economic analysis of liver transplantation. Gastroenterol Clin North Am 1993; 22:451–473.
3. Doyle HR, Marino IR, Jabbour N, et al. Early death or retransplantation in adults following orthotopic liver transplantation: can outcome be predicted? Transplantation 1994 (in press).
4. Maren AJ. Introduction to neural networks. *In* Maren A, Harston C, Pap R., eds. Handbook of Neural Computing Applications. San Diego, CA: Academic Press, 1990, pp 1–12.
5. Guyon L, Poujaud I, Personnaz L, et al. Comparing different neural network architectures for classifying handwritten digits. *In* Proceedings of the International Joint Conference on Neutral Networks. Volume II. New York: IEEE Press, 1989, pp 127–132.
6. Simpson PK. Neural networks for sonar signal processing. *In* Maren A, Harston C, Pap R, eds. Handbook of Neural Computing Applications. San Diego, CA: Academic Press, 1990, pp 319–335.
7. Smith WE, Barrett HH, Paxman RG. Reconstruction of objects from coded images by simulated annealing. Optics Letters 1983; 8: 199–201.
8. Miller WT, Sutton RS, Werbos RS. Neural Networks for Robotics and Control. Cambridge, MA: MIT Press, 1989.
9. Demeler B, Zhou GW. Neural network optimization for E. coli promoter prediction. Nucleic Acid Res 1991; 19:1593–1599.
10. Levine DS. Introduction to neural & cognitive modeling. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc, 1991.
11. Nadel L, Cooper LA, Culicover P, Harnish RM, eds. Neural Connections, Mental Computations. Cambridge, MA: MIT Press, 1989.
12. Baxt WG. Use of an artificial neural network for the diagnosis of myocardial infarction. Ann Intern Med 1991; 115:843–848.
13. Furlong JW, Dupuy ME, Heinsimer JA. Neural network analysis of serial cardiac enzyme data: a clinical application of artificial machine intelligence. Am J Clin Pathol 1991; 96:134–141.
14. Veselis RA, Reinsel R, Sommer A, Carlon G. Use of neural net-

work analysis to classify electroencephalographic patterns against depth of midazolam sedation in intensive care unit patients. J Clin Monit 1991; 7:259–267.
15. Ravdin PM, Clark GM. A practical application of neural network analysis for predicting outcome of individual breast cancer patients. Breast Cancer Res Treat 1992; 22:285–293.
16. Wu Y, Giger ML, Doi K, et al. Artificial neural networks in mammography: application to decision making in the diagnosis of breast cancer. Radiology 1993; 187:81–87.
17. Members of the American College of Chest Physicians/Society of Critical Care Medicine Consensus Conference Committee. American College of Chest Physicians/Society of Critical Care Medicine Consensus Conference: definitions for sepsis and organ failure and guidelines for the use of innovative therapies in sepsis. Crit Care Med 1992; 20:864–874.
18. Rumelhart DE, Hinton GE, Williams RL. Learning representations by back-propagating errors. Nature 1986; 323:533–536.
19. Maren AJ. Neural network structures: form follows function. *In* Maren A, Harston C, Pap R, eds. Handbook of Neural Computing Applications. San Diego, CA: Academic Press, 1990, pp 45–57.
20. Werbos P. Beyond regression: new tools for prediction and analysis in the behavioral sciences. Ph.D. Thesis to Harvard University Committee on Applied Mathematics. Cambridge, MA: Harvard University, 1974.
21. Rumelhart DE, Hinton GE, Williams RJ. Learning internal representations by error propagation. *In* Rumelhart DE, McClelland JL, eds. Parallel Distributed Processing: Explorations in the Microstructures of Cognition. Vol I. Cambridge, MA: MIT Press, 1986, pp 318–362.
22. Fahlman SE. Faster-learning variations on back-propagation: an empirical study. Proceedings of the 1988 Connectionist Models Summer School. Morgan Kaufmann, 1988, pp 38–51.
23. Hanley JA. Receiver operating characteristic (ROC) methodology: the state of the art. Crit Rev Diagn Imaging 1989; 29:307–35.
24. McCulloch WS, Pitts W. A logical calculus of the ideas immanent in nervous activity. Bull Math Biophysics 1943; 5:115–133.
25. Rosenblatt F. The perceptron: a probabilistic model for information storage and organization in the brain. Psychol Rev 1958; 65: 386–408.
26. Minsky M, Papert S. Perceptrons: an introduction to computational geometry. Cambridge, MA: MIT Press, 1969.
27. Hopfield JJ. Neural networks and physical systems with emergent collective computational abilities. Proc Natl Acad Sci 1982; 79: 2554–2558.
28. Miller AS, Blott BH, Hames TH. Review of neural network applications in medical imaging and signal processing. Med Biol Eng Comput 1992; 30:449–464.
29. Gross GW, Boone JM, Greco-Hunt V, Greenberg B. Neural networks in radiologic diagnosis. II. Interpretation of neonatal chest radiographs. Invest Radiol 1990; 25:1017–1023.
30. Silverman RH, Noetzel AS. Image processing and pattern recognition in ultrasonograms by backpropagation. Neural Networks 1990; 3:593–603.
31. Dassen WR, Mullereers R, Smeets J, et al. Self-learning neural networks in electrocardiography. J Electrocardiography 1990; 23: 200–202.
32. Edenbrandt L, Devine B, Macfarlane PW. Classification of electrocardiographic ST-T segments- human expert vs artificial neural network. European Heart J 1993; 14:464–468.
33. Mulsant BH. A neural network as an approach to clinical diagnosis. MD Computing 1990; 7:25–36.
34. Ebell MH. Artificial neural networks for predicting failure to survive following inhospital cardiopulmonary resuscitation. J Fam Pract 1993; 36:297–303.
35. Hornik K, Stinchcombe M, White H. Multilayer feedforward networks are universal approximators. Neural Networks 1989; 2:359–366.