Predicting Mortality after Coronary Artery Bypass Surgery:

What Do Artificial Neural Networks Learn?

JACK V. TU, MD, PhD, MILTON C. WEINSTEIN, PhD,
BARBARA J. MCNEIL, MD, PhD, C. DAVID NAYLOR, MD, DPhil, and
THE STEERING COMMITTEE OF THE CARDIAC CARE NETWORK OF ONTARIO*

Objective. To compare the abilities of artificial neural network and logistic regression models to predict the risk of in-hospital mortality after coronary artery bypass graft (CABG) surgery. *Methods.* Neural network and logistic regression models were developed using a training set of 4,782 patients undergoing CABG surgery in Ontario, Canada, in 1991, and they were validated in two test sets of 5,309 and 5,517 patients having CABG surgery in 1992 and 1993, respectively. *Results.* The probabilities predicted from a fully trained neural network were similar to those of a "saturated" regression model, with both models detecting all possible interactions in the training set and validating poorly in the two test sets. A second neural network was developed by cross-validating a network against a new set of data and terminating network training early to create a more generalizable model. A simple "main effects" regression model without any interaction terms was also developed. Both of these models validated well, with areas under the receiver operating characteristic curves of 0.78 and 0.77 ($p > 0.10$) in the 1993 test set. The predictions from the two models were very highly correlated ($r = 0.95$). *Conclusions.* Artificial neural networks and logistic regression models learn similar relationships between patient characteristics and mortality after CABG surgery. *Key words:* cardiac surgery; mortality; neural networks; logistic regression; ROC curves. **(Med Decis Making 1998;18:229–235)**

Recently, there has been widespread interest in using artificial neural networks (ANNs) for predicting clinical outcomes. Neural networks are pattern-recognition algorithms that are modeled after the biological structure of the human brain, and it has been suggested that they may offer some advantages over classic statistical approaches as predictive models for certain clinical problems.[1–7] ANNs are developed using an iterative training process in which training examples (e.g., previous patients who have undergone cardiac surgery) are repeatedly presented to a neural network. The weights of a network are gradually adjusted until the network "learns" the mathematical relationship between the predictor variables and the outcome of interest. Neural networks may theoretically offer greater pre-

Received December 16, 1996, from the Institute for Clinical Evaluative Sciences in Ontario (JVT, CDN) and the Clinical Epidemiology Unit and Division of General Internal Medicine, Department of Medicine, Sunnybrook Health Science Centre, University of Toronto, Toronto, Ontario, Canada (JVT, CDN); and the Department of Health Policy and Management, Harvard School of Public Health, Boston, Massachusetts (MCW) and Department of Health Care Policy, Harvard Medical School, Boston, Massachusetts (JVT, BJM). Revision accepted for publication June 12, 1997. Supported in part by a seed grant from the Harvard Center for Risk Analysis and by grants number HS08071 and HS08464 from the Agency for Health Care Policy and Research, Rockville, Maryland. Dr. Tu was supported by a Health Research Personnel Development Program Fellowship (04544) from the Ontario Ministry of Health. Dr. Naylor was supported by a Career Scientist Award (02377) from the Ontario Ministry of Health. The results and conclusions are those of the authors, and no official endorsement by the Ministry of Health is intended or should be inferred.

Address correspondence and reprint requests to Dr. Tu: Institute for Clinical Evaluative Sciences, G-106, 2075 Bayview Avenue, Toronto, Ontario, Canada, M4N 3M5.

*Steering Committee of the Cardiac Care Network of Ontario: Donald S. Beanlands, MD, University of Ottawa Heart Institute; Lorna Bickerton, BScN, University of Ottawa Heart Institute; Robert Chisholm, MD, St. Michael's Hospital, Toronto; Martin Goldbach, MD, Victoria Hospital, London; Vicki Kaminski, BScN, Sudbury Memorial Hospital; Jeffrey Lozon, MHA, St. Michael's Hospital, Toronto; Neil McKenzie, MB ChB, University Hospital, London; Barry J. Monaghan, BComm DHA, West Park Hospital, Toronto; Christopher D. Morgan, MD, Sunnybrook Health Science Centre, Toronto; John Pym, MB, BCh, Kingston General Hospital; Hugh Scully, MD, Toronto Hospital; B. William Shragge, MD, Hamilton General Hospital; James Swan, MD, Scarborough Centenary Health Centre.

dictive accuracy when important nonlinear or higher-order relationships that have not been detected by other forms of analysis exist in data sets.[1-4] However, relatively few head-to-head comparisons of logistic regression and neural network modeling techniques have been conducted by researchers in the field.

In this study, we developed ANN and logistic regression models to predict the risk of patient mortality following isolated coronary artery bypass graft (CABG) surgery (without concomitant valve surgery) in Ontario, Canada. Accurately identifying the risks of CABG surgery is an important clinical problem, and regression modeling techniques are frequently used for assessing surgical risk.[8-12] Risk-stratification models are useful both to clinicians counseling individual patients and to outcomes researchers comparing the risk-adjusted mortality rates of different hospitals and cardiac surgeons.[8-12] Since ANNs are often considered "black-box" prediction models, we sought to determine what neural networks actually learn about surgical risks during the training process.[13] In particular, we compared the predictions from ANN and logistic regression models, since greater predictive power has been cited as the major reason for using the more complex ANN modeling technique.

Table 1 • Patient Characteristics Used to Develop Predictive Models of In-hospital Mortality following Coronary Artery Bypass Graft (CABG) Surgery*

Age 65–74 years
Age ≥ 75 years
Female gender
Grade 2 left ventricular function (EF 35–50%)
Grade 3 left ventricular function (EF 20–34%)
Grade 4 left ventricular function (EF <20%)
Unknown left ventricular function
Urgent surgery
Emergency surgery
Previous CABG surgery
Left main disease (≥ 50% stenosis)
CCS Class 3 angina
CCS Class 4 angina
Recent myocardial infarction (<1 week)
Diabetes
Chronic obstructive lung disease
Peripheral vascular disease

*EF denotes ejection fraction; CCS, Canadian Cardiovascular Society angina class.

Table 2 • Summary of the Coronary Artery Bypass Graft (CABG) Surgery Data Sets*

	1991 Training Set	1992 Test Set	1993 Test Set
Patients	4,782	5,309	5,517
Unique covariate patterns	774	839	833
Patients who died	144	153	173
Unique covariate patterns with deaths	119	127	139
In-hospital mortality rate	3.01%	2.88%	3.14%

*The data sets were collected based on a fiscal year (April 1 of the calendar year to March 31 of the subsequent year).

Methods

DATA SOURCES

The data in this study come from the Cardiac Care Network (formerly the Provincial Adult Cardiac Care Network) of Ontario database, a large population-based cardiac surgery registry used to collect clinical data for all patients waiting for CABG surgery in Ontario. This clinical database was linked using unique patient identifiers to mortality and non-cardiac comorbidity information contained in the Canadian Institute for Health Information administrative database, as described elsewhere.[8,9] Eleven patient characteristics were examined as potential predictors of in-hospital mortality following CABG surgery (table 1), after identifying the most common surgical risk factors found in other studies.[8-12] Patient ages were initially grouped into five-year intervals and then these intervals were combined into three categories, < 65 years, 65–74 years, and ≥ 75 years, based on similar mortality rates within these age ranges.

In the current study, the database was divided into three data sets based on the fiscal years in which the data were collected (table 2). The training set included all patients who underwent isolated CABG surgery in Ontario in fiscal 1991 (April 1, 1991, to March 31, 1992) while the two test sets included all patients who had CABG surgery in fiscal 1992 (April 1, 1992, to March 31, 1993) and fiscal 1993 (April 1, 1993, to March 31, 1994). Since neural networks learn by recognizing patterns, the number of unique covariate patterns (i.e., unique combinations of patient characteristics) in each data set was determined, including the number of patterns associated with deaths each fiscal year.

ARTIFICIAL NEURAL NETWORK MODELS

Artificial neural network models were developed using the 1991 training set and were evaluated based on the two test sets using NeuroShell 2, a commercially available neural network simulator.[14] All models were developed using the back-propagation learning algorithm with a least-mean-squares-error objective function.[5,7,15] A logistic function was used in the output layer so that the network output would be similar to a probability.[16] Various neural network architectures were tried, and those yielding the best predictive performances were used for further comparisons with logistic regression models. The number of hidden nodes was varied between a lower limit of 1 and an upper limit of 100, and different types of activation functions in the hidden layer were tried (logistic, hyperbolic tangent, etc.).[14] An initial learning rate of 0.1 and a momentum term of 0.1 were chosen for developing these models.[5,15] Sensitivity analyses of these parameters were conducted to see whether they improved ANN performance. A detailed summary of all the models tried is available elsewhere.[17]

In this study, two different neural network models were developed. The first to be developed was a model designed for maximal predictive performance in the training set (i.e., the smallest error in the training set), hereafter referred to as ANN model 1. An analysis was conducted to determine the relationship between the predictions from this model and the actual probability of death for patients with each unique covariate pattern. Since this analysis showed that ANN model 1 overfit the data in the training set by memorizing each unique covariate pattern to the extent that the model was not generalizable, a second model, hereafter referred to as ANN model 2, was developed. This network was developed by periodically cross-validating the 1991 training set network on the 1992 test set during training and saving the network's weight configuration that performed best in the 1992 test set.[5] The performance of ANN model 2 in the 1993 test set should be considered the most valid test of this network's predictive performance because the 1992 test set was used in its development.

LOGISTIC REGRESSION MODELS

For comparison with ANN model 1, the equivalent of a "saturated" logistic regression model was created, with indicator variables used to represent every unique covariate pattern in the 1991 training set.[18] A saturated statistical model includes all possible interactions between the covariates. A main-effects logistic regression model was also developed for comparison with ANN model 2 by using indicator variables for the 11 covariates shown in table 1 (regression coefficients are not shown but are available from the first author).[17] A main-effects model includes the individual covariates but does not include any interaction terms. Predicted probabilities of patient deaths after CABG surgery were calculated using both the saturated model and the main-effects logistic regression model for all patients in the three data sets. The Stata statistical package was used for statistical analysis.[19]

COMPARISON OF NEURAL NETWORK AND LOGISTIC REGRESSION MODELS

The predicted probabilities of patient deaths after CABG surgery from ANN model 1 and the saturated logistic regression model were compared in the 1991 training set. Similarly, the predicted probabilities from the main-effects logistic regression model were compared with those from ANN model 2 in the 1993 test set, and Pearson's correlation coefficient between the predictions was determined. The areas under the receiver operating characteristic (ROC) curves were calculated for all four models in both the training set and the two test data sets.[20,21] The ROC curve is a measure of the discriminating ability of a model, with higher areas indicating better predictive ability. An ROC curve area of 1.00 indicates perfect predictive ability, while an area of 0.50 indicates predictability no better than chance.

Results

The overall predictive performances of the four CABG mortality predictive models are summarized in table 3. ANN model 1 had a very high area under the receiver operating characteristic curve of 0.88 in the 1991 training set but validated poorly in the 1992 and 1993 test sets, with areas under the ROC curves of 0.52 and 0.57, respectively. ANN model 1 was a network with an architecture containing 78 hidden nodes in one hidden layer (model not shown).[17] Analysis of the predicted probabilities from this model revealed that they were directly proportional to the mean probabilities of death for patients with the individual covariate patterns in the 1991 training set, a property associated with the predictions of saturated statistical models (see table 4 below).[18] In es-

Table 3 • Performances of Four Models for Predicting In-hospital Mortality after CABG Surgery

	Data Set	Area under the ROC Curve
Artificial neural network model 1*	1991 training set	0.88
	1992 test set	0.52
	1993 test set	0.57
Saturated logistic model	1991 training set	0.94
	1992 test set	0.46
	1993 test set	0.49
Artificial neural network model 2†	1991 training set	0.79
	1992 test set	0.77
	1993 test set	0.78
Main-effects logistic model	1991 training set	0.79
	1992 test set	0.76
	1993 test set	0.77

*1991 training-set error minimized.
†1992 test-set error minimized.

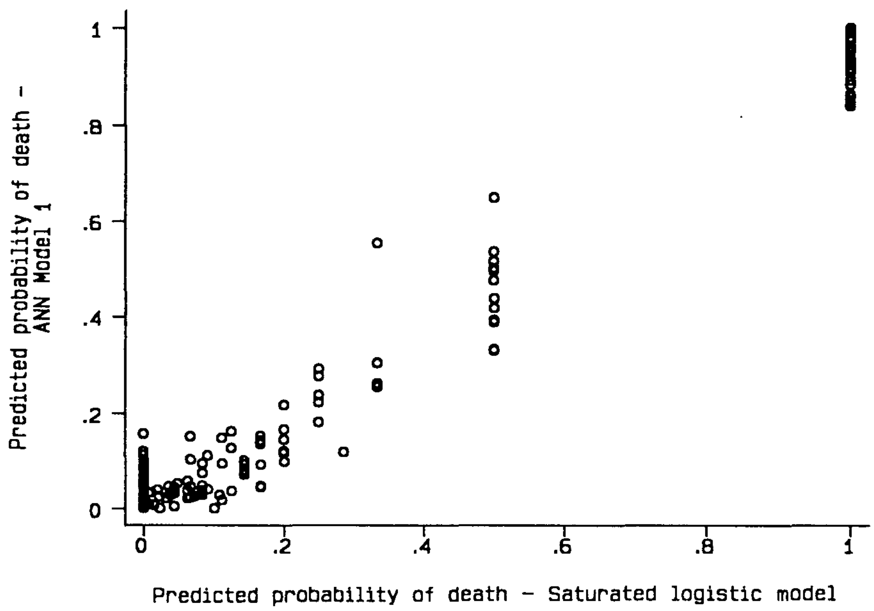


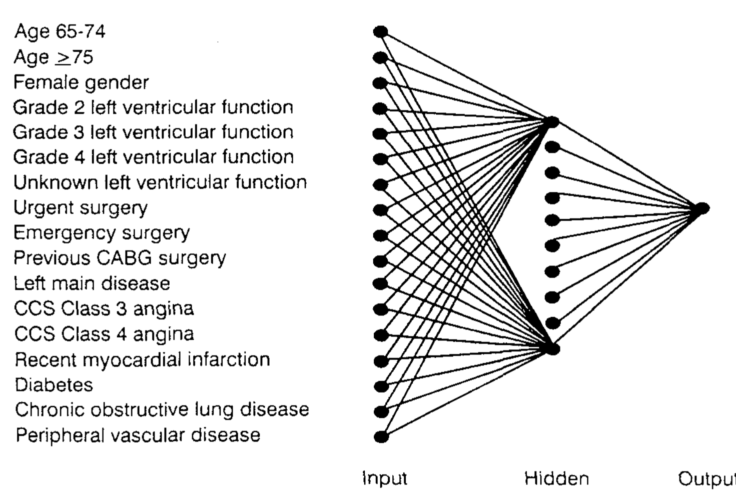
FIGURE 1. Predicted probabilities of patients dying after CABG surgery—ANN Model 1 versus saturated logistic regression model in the 1991 training set.

sence, this overfitted model memorized each specific pattern of clinical characteristics associated with deaths, reducing its generalizability to other years of data. The fully saturated logistic model had an even higher area under the ROC curve of 0.94 in the 1991 training set but also validated poorly in the 1992 and 1993 test sets. The predicted probabilities from ANN model 1 and the saturated logistic model are shown for all patients in the 1991 training set in figure 1.

ANN model 2 was the neural network model developed by intermittently cross-validating the 1991 training set network against the 1992 test set. This network was a model with one hidden layer containing ten hidden nodes, as shown in figure 2. In ANN model 2, each input variable to the network undergoes a nonlinear logistic transformation within each node in the hidden layer of the network before the predicted probability emerges from the output layer. Both ANN model 2 and the main-effects logistic regression model predicted mortality well in the 1993 test set, with areas under the ROC curve of 0.78 and 0.77 ($p > 0.10$).[20,21] Figure 3 shows a linear relationship between the predicted probabilities of death from these two models in the 1993 test set. The correlation between the predicted probabilities was very high ($r = 0.95$), suggesting that the two models learned similar relationships between patient characteristics and mortality after CABG surgery, although the predicted probabilities were spread over a wider range with the main-effects logistic regression model.

Predicted probabilities of patients dying after CABG surgery from the four prediction models are shown for some sample covariate patterns in table 4. This table demonstrates in detail that the predictions from ANN model 1 were directly proportional to the mean probabilities of death for patients with

FIGURE 2. Diagram of an (17 × 10 × 1) artificial neural network model (ANN Model 2) for predicting mortality after CABG surgery. Each circle represents a node, while each line represents a weight. Every node in each layer in connected to every node in the preceding and/or succeeding layers by a weight. Only the weights to the first and last nodes in the hidden layer are shown. A nonlinear logistic transformation is applied to the input variables at each node in the hidden layer and the output layer.



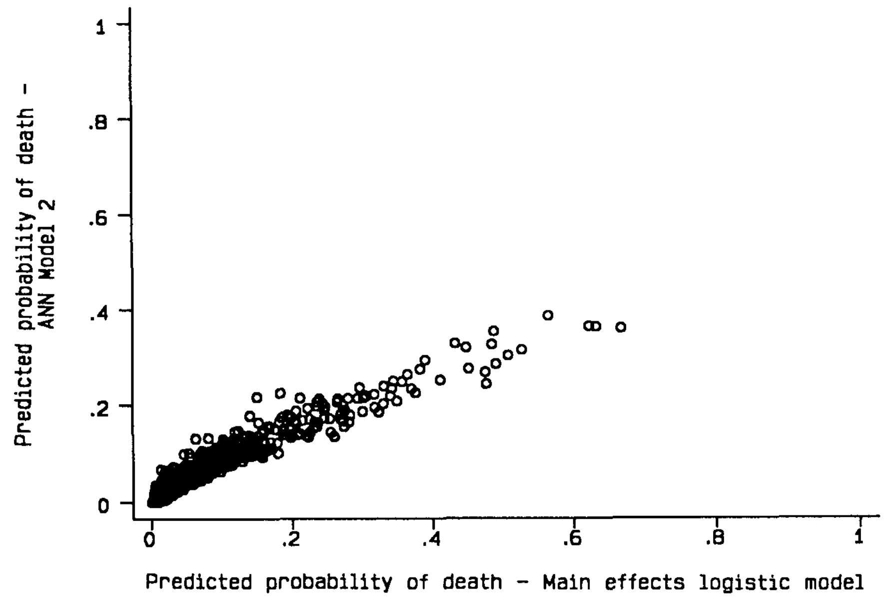


FIGURE 3. Predicted probabilities of patients dying after CABG surgery—ANN Model 2 versus main-effects logistic regression model in the 1993 test set.

individual covariate patterns (i.e., the predictions of a saturated regression model). This table also shows that the predictions from ANN model 2 were comparable in magnitude to the predictions of the main-effects logistic regression model. These results provide further evidence of the comparability of the predictions from these two techniques, and show that the two methods identify the same types of patients as being at the highest risk of death after CABG surgery.

Table 4 • Predicted Probabilities of Patients Dying after CABG Surgery from the Four Prediction Models for Some Sample Covariate Patterns in the 1991 Training Set

	Example							
	1 (*n* = 27)	2 (*n* = 18)	3 (*n* = 8)	4 (*n* = 4)	5 (*n* = 2)	6 (*n* = 2)	7 (*n* = 1)	8 (*n* = 1)
Mean probability of death	0	0	0.25	0.25	0.50	0.50	1.00	1.00
Predicted probability of death								
ANN* model 1	0	0.02	0.18	0.22	0.39	0.52	0.93	0.99
Saturated logistic model	0	0	0.25	0.25	0.50	0.50	1.00	1.00
ANN* model 2	0.02	0.01	0.07	0.09	0.04	0.06	0.13	0.18
Main-effects logistic model	0.01	0.01	0.10	0.13	0.05	0.07	0.13	0.28
Covariates in pattern								
Age 65–74 years					X	X		X
Age ≥75								
Female gender	X				X			X
Grade 2 LVF†			X				X	
Grade 3 LVF†		X		X				X
Grade 4 LVF†								
Unknown LVF†								
Urgent surgery	X							
Emergency surgery						X	X	X
Previous CABG‡ surgery			X	X				
Left main disease					X			X
CCS§ Class 3 angina		X		X	X			
CCS§ Class 4 angina	X		X			X	X	X
Recent myocardial infarction								
Diabetes		X						
COPD¶								
Peripheral vascular disease						X	X	

*Artificial neural network.
†Left ventricular function.
‡Coronary artery bypass graft.
§Canadian Cardiovascular Society.
¶Chronic obstructive pulmonary disease.

Discussion

In this study, we developed artificial neural network and logistic regression models for predicting the risk of patient death following isolated CABG surgery. We demonstrated that fully training a neural network (ANN model 1) leads a network to the equivalent of a "saturated" statistical state, with the network memorizing each unique cluster of clinical characteristics in the training set and its average mortality rate.[18] Differences between the training and test sets in the covariate patterns of patients who died caused this network to validate quite poorly in two test data sets. However, by cross-validating a neural network against a new set of data and terminating network training early in the training process, we were also able to develop a more generalizable model (ANN model 2) that predicted mortality as well as a main-effects logistic regression model in an independent test data set. Our study provides some novel insights into the statistical behavior of artificial neural networks, and demonstrates that neural network and logistic regression models learn similar relationships between patient characteristics and mortality after CABG surgery.

We embarked on this study with the hope that the neural network might uncover some previously unrecognized relationships between patient characteristics and mortality after CABG surgery, in light of some positive reports by other investigators.[1-4,22] However, the results of the study showed that the predicted probabilities from a trained network (ANN model 2) were very highly correlated with those of standard statistical models. Unlike regression models, which can clearly identify the patient characteristics that have the greatest influences on surgical risk, the "knowledge" acquired by the neural network during the training process is less obvious and is contained within the weights of the neural network. Our finding that the same types of patients were identified as high surgical risks by both the neural network and the regression model provides strong, albeit indirect, evidence that the knowledge about patient characteristics and surgical mortality acquired by the neural network is similar to that which we identified using regression analysis.

Our finding of similar levels of predictive performance with the neural network and the logistic regression models is not entirely unexpected. While some investigators have obtained impressive results with neural networks,[1-4] other studies suggest that neural networks offer small or no improvement over existing regression modeling techniques when head-to-head comparison studies have been conducted.[7] Neural networks may be expected to perform better when data sets contain very complex nonlinear relationships or many higher-order interactions between dependent and independent variables, but these types of data sets may be relatively infrequent in clinical medicine. Previous analyses of ours and other cardiac surgery data sets have shown that two-way interactions are usually not of major importance in assessing the risks of CABG surgery.[9-12] Furthermore, most of the variables that are risk factors for CABG mortality are binary variables (i.e., disease present or absent), so identifying nonlinear relationships are less likely to be important statistically. The only continuous variable in our study, age, was treated as a categorical variable so that we could perform analyses of individual covariate patterns.

Our study demonstrates that it is possible to mimic the pattern-recognition capabilities of neural networks by fitting a regression model with separate coefficients for each unique covariate pattern. However, this led to overfitting and the equivalent of a "saturated" statistical model.[18] To the best of our knowledge, the link between the neural network phenomenon of "overfitting" or memorizing all patterns in the training set and statistical "saturation" has not been previously recognized. For example, Doig and colleagues compared a fully-trained neural network (similar to our ANN model 1) and a main-effects logistic regression model in developing models to predict mortality in their intensive care unit.[23] They erroneously concluded that because the area under the ROC curve of the fully-trained neural network (0.9993) was superior to that of their logistic regression model (0.9259) in their training set neural networks were superior prediction models. However, the proper comparison should have been with a saturated regression model. A saturated regression model would have had the same or an even higher ROC curve area, since it represents the upper limit on any model's predictive performance in the training set.[18]

The relatively low prevalence (~3%) of deaths after CABG surgery meant that there were relatively few examples of each type of patient who died that the neural network could learn to recognize, even though there were thousands of patients in the training set. Since the covariate patterns of patients who die from CABG surgery vary from year to year, the neural network could only improve its recognition of these patients in the 1991 training set at the expense of its performance in the 1992 and 1993 test sets. These tradeoffs between predictive accuracy in the training set and the generalizability of a model explain why neural networks may not validate any better than regression models in certain data sets in spite of their ability to recognize individual covariate patterns. Neural networks may ultimately prove to be most useful in clinical scenarios requiring the recognition of small numbers of high-frequency covariate patterns (e.g., reading cervical smears, electrocardiograms).[24,25]

Our study has certain limitations. We restricted the study to using "back-propagation," the most popular neural-net–training algorithm in current use.[15] Newer algorithms are being developed by researchers in the field, and it is possible that some other training algorithm might have led to superior results for the neural network. Second, our study represents a comparison of neural networks and logistic regression in one data set only, and further comparisons of the two methods should be conducted by researchers using other data sets. In particular, studies using data sets with more continuous variables or more frequent adverse outcomes could yield different results. Third, we restricted the input variables in our study to those that have previously been shown to be important predictors of CABG mortality. Some investigators have suggested that neural networks can identify variables that are not important outcome predictors by traditional statistical methods.[22]

In conclusion, we have demonstrated a remarkable similarity in the abilities of artificial neural network and logistic regression models to predict in-hospital mortality after CABG surgery. Neural networks and logistic regression models appear to learn similar relationships between patient characteristics and mortality after CABG surgery, but this knowledge is modeled in a less transparent manner in a neural network. Although artificial neural networks are an interesting alternative to classic statistical methods for predicting surgical outcomes, we did not find any significant predictive advantage to using them for predicting mortality after CABG surgery. For our ongoing analyses of cardiac surgical outcomes in Ontario, we plan to rely on logistic regression and other more conventional statistical methods to assess the outcomes of CABG surgery.[8,9]

The authors thank Carl N. Morris, PhD, for his helpful advice on this project, and Keyi Wu, MSc, for assistance with computer programming. They also thank the cardiovascular practitioners, nurses, and registry personnel who make up the Cardiac Care Network of Ontario for participating in the study.

References

1. Baxt WG. Use of an artificial neural network for the diagnosis of myocardial infarction. Ann Intern Med. 1991;115:843–8.
2. Baxt WG, Skora J. Prospective validation of artificial neural network trained to identify acute myocardial infarction. Lancet. 1996;347:12–5.
3. Buchman TG, Kubos KL, Seidler AJ, Siegforth MJ. A comparison of statistical and connectionist models for the prediction of chronicity in a surgical intensive care unit. Crit Care Med. 1994;22:750–62.
4. Dybowski R, Weller P, Chang R, Gant V. Prediction of outcome in critically ill patients using artificial neural network synthesized by genetic algorithm. Lancet. 1996;347:1146–50.
5. Tu JV. Advantages and disadvantages of using artificial neural networks versus logistic regression for predicting medical outcomes. J Clin Epidemiol. 1996;49:1225–31.
6. Cross SS, Harrison RF, Kennedy RL. Introduction to neural networks. Lancet. 1995;346:1075–9.
7. Penny W, Frost D. Neural networks in clinical medicine. Med Decis Making. 1996;16:386–98.
8. Tu JV, Naylor CD. Steering Committee of the Provincial Adult Cardiac Care Network of Ontario. Coronary artery bypass mortality rates in Ontario: a Canadian approach to quality assurance in cardiac surgery. Circulation. 1996;94:2429–33.
9. Tu JV, Jaglal SB, Naylor CD. Steering Committee of the Provincial Adult Cardiac Care Network of Ontario. Multicenter validation of a risk index for mortality, intensive care unit stay, and overall hospital of stay after cardiac surgery. Circulation. 1995;91:677–84.
10. Higgins TL, Estafanous FG, Loop FD, Beck GJ, Blum JM, Paranandi L. Stratification of morbidity and mortality outcome by preoperative risk factors in coronary artery bypass patients: a clinical severity score. JAMA. 1992;267:2344–8.
11. O'Connor GT, Plume SK, Olmstead EM, et al. Northern New England Cardiovascular Disease Study Group. Multivariate prediction of in-hospital mortality associated with coronary artery bypass graft surgery. Circulation. 1992;85:2110–8.
12. Hannan EL, Kolburn H Jr, Racz M, Shields E, Chassin MR. Improving the outcomes of coronary artery bypass surgery in New York State. JAMA. 1994;271:761–6.
13. Hart A, Wyatt J. Evaluating black-boxes as medical decision aids: issues arising from a study of neural networks. Med Inform. 1990;15:229–36.
14. Neuroshell 2. Ward Systems Group, Inc., Frederick, MD.
15. Rumelhart DE, Hinton GE, Williams RJ. Learning representations by back-propagating errors. Nature. 1986;323:533–6.
16. Richard MD, Lippman RP. Neural network classifiers estimate Bayesian a posteriori probabilities. Neural Computation. 1991;3:461–3.
17. Tu JV. Quality of cardiac surgical care in Ontario, Canada [dissertation]. Cambridge, MA: Harvard University, 1996.
18. Shwartz M, Ash AS. Evaluating the performance of risk-adjustment methods: continuous measures. In: Iezzoni LI (ed). Risk Adjustment for Measuring Health Care Outcomes. Ann Arbor, MI: Health Administration Press, 1994:287–311.
19. Stata Release 4.0. College Station, TX, 1995.
20. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology. 1982;143:29–36.
21. Hanley JA, McNeil BJ. A method of comparing the areas under receiver operating characteristic curves derived from the same cases. Radiology. 1983;148:839–43.
22. Baxt WG. Analysis of the clinical variables driving decision in an artificial neural network trained to identify the presence of myocardial infarction. Ann Emerg Med. 1992;21:1439–44.
23. Doig GS, Inman KJ, Sibbald WJ, Martin CM, Robertson JM. Modeling mortality in the intensive care unit: comparing the performance of a back-propagation, associative-learning neural network with multivariate logistic regression. In: Proceedings of the Seventeenth Annual Symposium on Computer Applications in Medical Care. Washington, DC: McGraw–Hill, 1993; pp 361–5.
24. Koss LG, Lin E, Schreiber K, Elgert P, Mango L. Evaluation of the PAPNET cytologic screening system for quality control of cervical smears. Am J Clin Pathol. 1994;101:220–9.
25. Heden B, Ohlsson M, Rittner R, et al. Agreement between artificial neural networks and experienced electrocardiographer on electrocardiographic diagnosis of healed myocardial infarction. J Am Coll Cardiol. 1996;28:1012–6.