Published in final edited form as:

Transplantation. 2008 November 27; 86(10): 1349–1350. doi:10.1097/TP.0b013e31818b2417.

For an Always Promising Transplant Prediction, Call ANN

By David W. Gjertson^{1,2,*} and Bill D. Clark¹

¹ *UCLA Immunogenetics Center, Department of Pathology & Laboratory Medicine, David Geffen School of Medicine at UCLA*

² *Department of Biostatistics, UCLA School of Public Health*

Keywords

Renal transplantation; Predicting graft survival; Artificial neural networks; Nomograms

With apologies to Sherlock Holmes, “You can never foretell when any one man’s kidney transplant will fail, but you can say with precision when an average number will fail. ... So says the statistician.”

Predicting the fate of every clinical transplant is a notable, unsolved exercise. In this issue of *Transplantation*, Akl, Mostafa and Ghoneim attempt this level of prognostication, describing two potential methods – an artificial neural network (ANN) and a scoring nomogram calibrated from Cox regression coefficients (1). The authors constructed their models based on a cohort of 1900 recipients transplanted with living-donor kidneys between 1976–2007 and used 11 significant factors identified from univariate analyses. The study was designed to predict 5-year survival using cases with complete data (81 with missing data were dropped) divided into training (n=1500) and test (n=319) sets for building and validating models, respectively. The ANN exhibited better performance characteristics than the nomogram. For example, the area under the ROC curve (a measure of predictive accuracy) was significantly improved in the test set: 0.88 for the ANN versus 0.72 for the nomogram. The authors concluded that the ANN model was superior. Their ANN is a promising prototype but is unsupported by relevant clinical implementation.

The ANN used in this study was developed from commercial software (SPSS Inc.) utilizing a feed-forward, back-propagation architecture (2). The algorithm is often described as a decision-making process functioning like the human brain (3). The ANN forms a mapping from input to output nodes (simulated neurons) by extracting features from input patterns, assigning them weights, summing weights with activation functions, and propagating decisions to output nodes if activation thresholds are exceeded.

Typical networks are organized into 3 layers of computational units (nodes) in which input/output layers are linked by hidden layers of nodes. Subject factors determine the number of input units, and the classification complexity determines the number of output units. The number of hidden units is determined by trial and error (training). Common routines start with one hidden unit and assign small arbitrary weights to all nodal connections. The network is fed sample data with known outcomes, and an error term is calculated via differences between known and predicted outputs. Learning consists of adjusting weights by backward pass of errors through the connections to network nodes in response to input data. Hidden units are

*Corresponding Author and Address: David W. Gjertson, PhD, UCLA Immunogenetics Center, David Geffen School of Medicine at UCLA, 1000 Veteran Avenue, Los Angeles, California, 90095, USA, E-mail: gjertson@ucla.edu, Tel: 001-310-206-0255, Fax: 001-310-206-3216.

added to achieve minimum error criteria, while constraining the number to promote generalization of input patterns and prevent overfitting (memorization). Interconnection density determines the network's ability to correctly discriminate outcomes.

In statistical parlance, ANN models are a form of non-linear discriminant analysis with input units, weights and activation functions resembling covariates, coefficients and generalized additive models, respectively. Thus, it is not surprising that ANN applications are undermined by similar limitations and misuses afflicting conventional discriminant analysis. Schwarzer et al. identified four frequent mistakes when applying ANNs (4):

- Overfitting models by training large, multi-layer networks with small data sets.
- Neglecting traditional statistical methods due to inadequate benchmarks or lack of significance testing.
- Applying naïve approaches to survival data, sometimes ignoring censorship.
- Claiming overly optimistic generalization properties.

Here, the authors are to be congratulated for taking steps to guard against these problems. To avoid overfitting, the ANN was restricted to one hidden layer, and the number of hidden nodes was controlled by reasonable stopping criteria. Also, the ratio of number of observations in the training set (1500) to number of parameters in the model (361) was greater than 2, a recommended guideline (4). The ANN model was tested against a nomogram built on a standard statistical approach (Cox regression). The nomogram was not discarded altogether but credited for easy interpretation. Albeit vague, the authors described a survival neural network that imputes outcomes for patients whose follow-up was less than 5 years (2). The imputation strategy appeared to be sufficient since calibration points between average predicted (ANN) and observed (Kaplan-Meier) 5-year survival rates for various patient groups produced virtually a 45° line.

On the other hand, several aspects of the current study are troublesome, especially with regard to generalizing results. First, the testing group was taken from a specific time period, which may cause bias when validating the model. Normally, testing and training groups are derived randomly. Second, only significant univariate variables were incorporated as input units, but in multivariate cases, insignificant univariate variables sometimes become relevant confounders or effect modifiers. Since ANNs are touted as having the ability to select those items most important in performing classifications, prior variable selection seems unnecessary. Third, the authors failed to describe procedures for preprocessing or normalizing data entering the ANN. This is an important step that usually requires a floating-point value between 0 and 1 to be assigned for each input node, with special consideration for missing values. Since the authors didn't allow any data with missing values to enter their study, the constructed ANN would be impractical for 'real world' use. Fourth, their static feed-forward ANN cannot re-tune to accommodate new conditions should the statistical make-up of the population change. To test whether their model is detecting important data patterns, they should try validating with records from other centers.

ANNs have certainly helped to predict a wide range of outcomes in health and medical research; however, moving ANN models into clinical settings has met with poor success due to the lack of appropriate comparative trials, ones which assess the overall decision making capabilities of an intervention involving ANNs (5). Appropriate trials are hampered by facts that medical data often exhibit poor signal-to-noise ratios, standard protocols vary with center, and patient populations drift (4). Transplant data suffers these drawbacks, too, i.e., individual transplant factors explain little of the overall variation in outcome, numerous protocols exist for listing and maintaining patients, and recipient/donor populations differ in many characteristics such

as race, age and gene polymorphisms. Consequently, we predict that Akl and others will experience very little success in translating their ANN prognostications into clinical practice.

References

1. Akl A, Mostafa A, Ghoneim M. Prediction of graft survival of living donor kidney transplantation: Nomograms or Artificial Neural Networks. *Transplantation* 2008;xxx:xxx-xxx.
2. Ripley RM, Harris AL, Tarassenko L. Non-linear survival analysis using neural networks. *Stat Med* 2004;23:825-42. [PubMed: 14981677]
3. Cross SS, Harrison RF, Kennedy RL. Introduction to neural networks. *Lancet* 1995;346:1075-79. [PubMed: 7564791]
4. Schwarzer G, Vach W, Schumacher M. On the misuses of artificial neural networks for prognostic and diagnostic classification in oncology. *Stat Med* 2000;19:541-61. [PubMed: 10694735]
5. Lisboa PJG. A review of evidence of health benefit from artificial neural networks in medical intervention. *Neural Networks* 2002;15:11-39. [PubMed: 11958484]