Assessment of a simple artificial neural network for predicting residual neuromuscular block

J. G. Laffey, É. Tobin, J. F. Boylan and A. J. McShane*

Department of Anaesthesia, Intensive Care and Pain Medicine, St Vincent’s University Hospital, Dublin, Ireland

*Corresponding author: Department of Anaesthesia, Intensive Care and Pain Medicine, St Vincent’s University Hospital, Elm Park, Dublin 4, Ireland. E-mail: l.mcnicholas@st-vincents.ie

Background. Postoperative residual curarization (PORC) after surgery is common and its detection has a high error rate. Artificial neural networks are being used increasingly to examine complex data. We hypothesized that a neural network would enhance prediction of PORC.

Methods. In 40 previously reported patients, neuromuscular function, neuromuscular block/antagonist usage and time intervals were recorded throughout anaesthesia until tracheal extubation by an observer uninvolved in patient care. PORC was defined as significant ‘fade’ (train of four <0.7) at extubation. Neuromuscular function was classified as PORC (value=1) or no PORC (value=0). A back-propagation neural network was trained to assign similar values (0, 1) for prediction of PORC, by examining the impact of (i) the degree of spontaneous recovery at reversal, and (ii) the time since pharmacological reversal, using the jackknife method. Successful prediction was defined as attainment of a predicted value within 0.2 of the target value.

Results. Twenty-six patients (65%) had PORC at tracheal extubation. Clinical detection of PORC had a sensitivity of 0 and specificity of 1, with an indeterminate positive predictive value and a negative predictive value of 0.35. Using the artificial neural network, one patient with residual block and one with adequate neuromuscular function were incorrectly classified during the test phase, with no indeterminate predictions, giving an artificial neural network sensitivity of 0.96 (χ²=44, P<0.001) and specificity of 0.92 (P=1), with a positive predictive value of 0.96 and a negative predictive value of 0.93 (χ²=12, P<0.001).

Conclusions. Neural network-based prediction, using readily available clinical measurements, is significantly better than human judgement in predicting recovery of neuromuscular function.

Br J Anaesth 2003; 90: 48–52

Keywords: measurement techniques, train-of-four; model, artificial neural network; neuromuscular block, atracurium

Accepted for publication: August 30, 2002

Residual neuromuscular block after surgery may be a significant problem, even after the use of medium- or short-acting agents. It is uncertain whether peripheral nerve stimulator (PNS) use can reduce the incidence of clinically significant postoperative residual curarization (PORC). Human error in assessing PNS data is very common, possibly stemming in part from perceptual limitations. Errors of judgement in anaesthetic practice may also result from pressure of work. A train-of-four value greater than 0.7 has been suggested as a minimum criterion for safe tracheal extubation, as lower values may be associated with impaired airway protection. Even higher train-of-four threshold values have been proposed, although they cannot be reliably assessed by the naked eye. Acceleromyographic techniques, while gaining in popularity, may not agree reliably with mechanomyographically derived data, limiting their validity. Transduced or electromyographic monitoring systems are unlikely to enter clinical practice in the immediate future, suggesting that the assessment of neuromuscular block will remain semiquantitative.

This work was presented in part at the National Scientific Meeting of the Royal College of Physicians in Ireland, Dublin, March 1998, and at the International Anesthesia Research Society, Honolulu, Hawaii, March 2000.
Several variables predict successful reversal of neuromuscular block. However, the failure rates observed suggest that these are imperfect and insensitive in practice. Artificial neural networks consist of computer software designed to mimic multiple inputs and non-linear interactions, and are being increasingly used to examine complex data. We tested the hypothesis that a neural network-based analysis would enhance the prediction of PORC when compared with human decision-making.

Patients and methods
We retrospectively examined data on neuromuscular function during and after general anaesthesia, obtained during a previous study. After hospital ethical approval and individual informed consent, 40 patients were recruited. Data were collected by an observer (E.T) uninvolved in clinical care. The transduced twitch height and train-of-four values were collected from immediately after induction of anaesthesia and before atracurium administration up to the time of tracheal extubation. Atracurium dosages, antagonists and time intervals from induction through antagonism to tracheal extubation were recorded. Anaesthetic care was provided by residents in training under the direction of a consultant anaesthetist; care providers were blinded to the data being collected and clinical anaesthetic management was not influenced by research personnel. For the purposes of the study, a train-of-four value of 0.7 or less at the time of antagonism; and (ii) the time elapsed from neostigmine administration until tracheal extubation. The single-node output layer delivered a value of 1 or 0 corresponding to the presence or absence, respectively, of residual neuromuscular block in the training data set. The number of training cycles and the learning rate (viz., the degree of weighting adjustment between cycles) was preset by the investigators. Successful prediction during the training and test phases was defined as a performance within a tolerance of 0.2, where tolerance was defined as a difference between observed and predicted output (0, 1) of 0.2 or less. Thus, successful prediction of residual block corresponded to a predicted output value of 0.8–1.0 for an observed output value of 1, while successful prediction of adequate reversal corresponded to a predicted output value of 0.0–0.2 for an observed output value of 0. Misclassification was defined as a predicted output value of 0.8–1.0 for an observed output value of 0 (i.e. an erroneous prediction of inadequate neuromuscular function), or predicted output value of 0.0–0.2 for an observed output value of 1 (i.e. an erroneous prediction of adequate neuromuscular function). Other predicted values (0.21–0.79) were considered indeterminate. During preliminary data analysis, a broad range of learning rates were chosen, in all of which the model reached optimal predictive performance within 5000 cycles, each cycle representing a scan through the data with a training update. We chose our two input variables on the basis of previous work indicating that the most important determinants of successful recovery of neuromuscular function are the degree of spontaneous neuromuscular recovery before antagonism, and the time elapsed since administration of pharmacological antagonism. To ensure validity of analysis (specifically, to prevent overlearning, in which data points are memorized rather than patterns being detected), the network was trained a total of 40 times using...
these variables, using the ‘leave-out-\( k \)' or jackknife method, which trained the neural network on all data except for a group (in this case, one patient) that was subsequently used to test the network’s conclusions. A pass through the data set using every individual in turn as a single test case allowed comparison of the predictive powers of the training and test phases. At the start of each new training period, previous weightings were erased, so that test cases were evaluated using fresh training data.

For diagnostic and predictive purposes, ‘positive’ was defined as the presence of residual neuromuscular block, and ‘negative’ was defined as the absence of residual neuromuscular block. Neural network sensitivity and specificity data were derived by sequentially using each of the 40 patients as the test case. Human and neural network performances were compared using exact probability test as appropriate.

**Table 1** Comparison of diagnostic performances of clinical diagnosis and neural network-based (ANN) prediction

<table>
<thead>
<tr>
<th></th>
<th>PORC(^+)</th>
<th>PORC(^-)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Incidence</td>
<td>Sensitivity</td>
</tr>
<tr>
<td>Clinical</td>
<td>26/40</td>
<td>0/26</td>
</tr>
<tr>
<td>ANN</td>
<td>26/40</td>
<td>25/26</td>
</tr>
<tr>
<td></td>
<td>14/40</td>
<td>14/14</td>
</tr>
<tr>
<td></td>
<td>14/40</td>
<td>13/14</td>
</tr>
</tbody>
</table>

PORC=postoperative residual curarization; PORC\(^+\)=presence of PORC; PORC\(^-\)=absence of PORC.

Discussion

In this small test group, with a high incidence of residual neuromuscular block, a simple neural network was able to predict the likelihood of normal or abnormal neuromuscular function at the time of the decision to perform tracheal extubation with greater accuracy than that of clinical assessment. This suggests that, even with limited, imprecise data, artificial neural network-based prediction of simple drug pharmacodynamic relationships can equal or outperform human assessment in this setting.

Increasing processor speeds and computer power have moved neural networks from an esoteric topic to one where they offer practical uses in a wide range of fields. At their most basic they can be used as cheap, small programs (essentially macro add-ins to commercial spreadsheets). Successful medical uses of neural networks include their use in pharmacokinetic/pharmacodynamic prediction (antibiotic peak and trough levels)\(^{13}\) and enhancement of diagnostic skills. In the clinical arena, the emergency room diagnosis of myocardial infarction\(^{14}\) and radiologists’ diagnosis of pulmonary embolism\(^{15}\) have been studied with favourable results. A neural network for detecting oesophageal intubation has been described.\(^{16}\) More recently, neural network-based prognostication in critical care has been shown to be superior to a conventional statistical approach.\(^{17}\) Enhanced performance by neural networks relative to humans is related to (i) more appropriate weighting given by the network to common diagnostic factors, and (ii) empirical use by networks of diagnostic rules of thumb not previously utilized. In this study, the diagnostic variables used by the software were deliberately limited to two of the commonest ones used in clinical practice, instead of allowing all possible information to be
Neural network prediction of neuromuscular block

used. Although one patient with residual block was consistently misclassified by the network, less common predictors of incomplete reversal are beyond the scope of this study. Our findings suggest that their assessment would require much larger patient numbers. For example, assuming a successful prediction rate of 90% by the network using only the two variables employed in this study, in order to have an 80% chance of detecting a further improvement in performance to, say, 95%, we would have needed about 240 patients.

Neural networks have several limitations. A major theoretical concern is the ‘black box’ nature of their output, i.e. conclusions are generated without explanations. This has led to claims that they lack rigour, and a reluctance to accept their results. This criticism implies that conclusions should follow from hypotheses supported by data and rejects the role of pattern recognition in decision making. However, the success of neural networks in medical decision-making, including a performance equal to that of radiologists in diagnosis of pulmonary embolism,15 and superior to that of emergency physicians in the diagnosis of myocardial infarction,16 suggests that they are competent at the very least. Black box concerns may be overcome by sensitivity analyses, where the effects of variables are assessed by their inclusion or exclusion. Our a priori choice of a small number of predictor variables in this study also addresses this criticism.

Neural networks may be difficult to validate in practice. If settings, especially learning rate, are not chosen carefully during the training phase, overlearning may occur. In overlearning, the network memorizes the data points and gives correct answers without recognizing patterns at all. As in this case, the problem is addressed by using low values for learning rate at the expense of decreased speed, and/or using separate training and test sets. The jackknife technique used during this study successfully prevented overlearning, as demonstrated by similar performances for the training and test phases, without reducing success rates. The sensitivity and specificity of any diagnostic or predictive technique will vary as a function of the cut-off (tolerance) points used to determine agreement. The success rate seen with tolerance values of 0.2 indicates that, in 95% of cases, the network correctly predicted adequate or inadequate return of neuromuscular function.

The high success rate with this model may be partly related to an unusually high incidence of undetected residual block in the training data set. The presumed failure to make a clinical diagnosis of PORC in any patient limited our ability to make a comparison of indices such as sensitivity and positive predictive value; however, the negative predictive values suggest that, compared with clinical care providers, the artificial neural network was much more effective in predicting recovery of adequate neuromuscular block. Even board-certified anaesthetists may fail to diagnose residual relaxant effects, as evidenced by a 25% requirement for unscheduled antagonism in adults receiving mivacurium without pharmacological antagonism in a clinical trial setting.18 It is possible that trainees have even less capacity to assess the speed and degree of pharmacological antagonism and that this could be improved significantly. The pilot nature of these data and the fact that they provide a single point estimate of the prevalence of PORC in a small number of patients make it premature to suggest that neural networks be used for real-time decision-making in the antagonism of neuromuscular block.

Neural networks are limited by the quality of their data and may need to be retrained periodically if performance changes over time, but this may be less applicable in the assessment of drug pharmacokinetics and pharmacodynamics. In predicting peak and trough antibiotic levels, network-based prediction has already been shown to be as good as or better than standard population-based pharmacokinetic modelling.13 For optimal assessment of network-based prediction, further validation would be required in the form of a prospective comparison with clinical practice and conventional statistical analysis. Also, higher train-of-four values (of the order of 0.9) are required to ensure full airway competence10 and minimize subjective discomfort.2 Because such data are not clinically detectable with routine monitoring, neural network techniques may play a role in enhancing prediction of greater degrees of neuromuscular recovery. This area merits further study.

We deliberately chose simple variables as predictors of residual block. Other data collected, such as absolute values for twitch height depression, were excluded from analysis, because they are not routinely measured in practice. In the present study, the predictive ability of the neural network was so strong using two simple indices that incorporating extra variables could not have enhanced performance with this sample size. This suggests that most of the variation seen in neuromuscular antagonism is attributable to these two factors. The difference in negative predictive values obtained using the two methods suggests that routinely collected variables are adequate in most patients for assessment of neuromuscular block when emphasis is placed on specificity (i.e. priority is given to exclusion rather than detection of PORC). Although our data are limited to the assessment of a single clinical decision (i.e. whether to extubate) in a small series of patients, our findings suggest that human factors are a major component in the erroneous diagnosis of adequate neuromuscular antagonism.

In conclusion, using simple, routinely measured variables, a neural network-based prediction is useful in estimating the likelihood of clinically significant residual neuromuscular block at the time of tracheal extubation, with a significant performance improvement over that of the anaesthetic trainees from whose practice the data were derived. Other predictive problems in anaesthesia where low sensitivity or non-linear interactions are a limiting factor may be amenable to neural network-based analysis.
Acknowledgement
É. Tobin and A. J. McShane were supported in part by the Health Research Board (Ireland).

References
2. Kopman AF, Yee PS, Neuman GG. Relationship of the train-of-four fade ratio to clinical signs and symptoms of residual paralysis in awake volunteers. Anesthesiology 1997; 86: 765–71