British Journal of Anaesthesia **98 (1)**: 60–5 (2007)
doi:10.1093/bja/ael282 Advance Access publication October 25, 2006

BJA

Predicting postoperative nausea and vomiting with the application of an artificial neural network

S. Y. Peng[1] *, K. C. Wu[2], J. J. Wang[2], J. H. Chuang[3], S. K. Peng[1] and Y. H. Lai[2]

[1]*Department of Anesthesiology, Taichung Veterans General Hospital, Taiwan, Republic of China.*
[2]*Department of Anesthesiology, Shin Kong Wu Ho-Su Memorial Hospital, Taiwan, Republic of China.*
[3]*Center for Disease Control and Institute of Health Informatics and Decision Making, School of Medicine, National Yang-Ming University, Taiwan, Republic of China*

**Corresponding author: No. 108-3, Section 2, Fuxing Road, South District, Taichung City 402, Taiwan, Republic of China. E-mail: shihyen.peng@gmail.com*

Background. Several medications have proved to be useful in preventing postoperative nausea and vomiting (PONV). However, routine antiemetic prophylaxis is not cost-effective. We evaluated the accuracy and discriminating power of an artificial neural network (ANN) to predict PONV.

Methods. We analysed data from 1086 in-patients who underwent various surgical procedures under general anaesthesia without antiemetic prophylaxis. Predictors used for ANN training were selected by computing the value of χ^2 statistic and information gain with respect to PONV. The configuration of the ANN was chosen by using a software tool. Then the training of the ANN was performed based on data from a training set (n=656). Testing validation was performed with the remaining patients (n=430) whose outcome regarding PONV was unknown to the ANN. Area under the receiver operating characteristic (ROC) curves were used to quantify predictive performance. ANN performance was compared with those of the Naïve Bayesian classifier model, logistic regression model, simplified Apfel score and Koivuranta score.

Results. ANN accuracy was 83.3%, sensitivity 77.9% and specificity 85.0% in predicting PONV. The areas under the ROC curve follow: ANN, 0.814 (0.774–0.850); Naïve Bayesian classifier, 0.570 (0.522–0.617); logistic regression, 0.669 (0.623–0.714); Koivuranta score, 0.626 (0.578–0.672); simplified Apfel score, 0.624 (0.576–0.670). ANN discriminatory power was superior to those of the other predicting models (P<0.05).

Conclusions. The ANN provided the best predictive performance among all tested models.

Br J Anaesth 2007; **98**: 60–5

Keywords: measurement techniques, artificial neural network, Naïve Bayesian classifier, Koivuranta score; PONV

Accepted for publication: August 23, 2006

Introduction

Despite introduction of new anaesthetic agents with reduced emetogenicity, the incidence of postoperative nausea and vomiting (PONV) remains 20–30%.[1–4] Aetiology is multifactorial, including anaesthetic and analgesic factors, type of surgery and patient-related factors.[4 5] PONV can cause suture rupture, wound dehiscence, pulmonary aspiration, bleeding, dehydration and electrolyte disturbance.[1 6 7] Some patients regard PONV as more distressing than postoperative pain.[8] Although antiemetic prophylaxis can be helpful,[9] routine efforts are generally contraindicated by cost and side-effects of antiemetics. So far numerous predictive scores and models have been developed, but they have limited discriminating power.[10–12]

An artificial neural network (ANN) is an abstract computational model of the human brain. ANNs can learn from experiential knowledge expressed through interunit connection strengths, and they can make such knowledge available for use.[13] In the studies of Eberhart and Traeger, ANNs were used to predict PONV in patients who received total i.v. anaesthesia (TIVA).[14 15] In this study, we applied an ANN to predict PONV in adult inpatients who received general anaesthesia. Furthermore, we compared the performance of our ANN to those of the Naïve Bayesian classifier,

logistic regression, Koivuranta score and simplified Apfel score.

Materials and methods

After receiving approval from the Ethics Committee, we prospectively studied 1086 adult in-patients who received surgery under general anaesthesia during a 2 month period at the Shin Kong Wu Ho-Su Memorial Hospital. Because the study entailed no interventions in patient care, written informed consent was not required by the Ethics Committee. In-patients who underwent elective surgery under general anaesthesia were included. We excluded paediatric patients (age <18 yr), patients who needed postoperative ventilator support or intensive care, patients who received general anaesthesia combined with spinal or epidural anaesthesia, patients who received preoperative or intraoperative antiemetic drugs, and patients who received prophylactic antiemetic drugs within 24 h after surgery. Patients who experienced dry heaving, a feeling of nausea, retching or vomiting during the period of postoperative 24 h were recognized as PONV. Most physicians of this hospital are used to treating PONV after its occurrence. Patients who underwent ophthalmic procedures were excluded because they were routinely given prophylactic antiemetics.

Preoperative patient characteristics and intraoperative variables were documented on standardized adverse-outcome check-off forms that were completed by nursing staff. All patients received general anaesthesia and standardized monitoring of pulse rate, blood pressure, pulse oximetry, level of consciousness, ventilatory frequency and temperature. General anaesthesia was provided by different anaesthesiologists without any restrictions. Postoperative interviews were conducted 24 h after surgery by the nursing staff. Patient characteristics, ASA status, duration of anaesthesia and surgical procedure were documented in the interview record. Patients who underwent surgery under general anaesthesia during the first month of the study period (n=656) were used for training the ANN, and the rest of the study population (n=430) were used for validation with the ANN, Naïve Bayesian classifier, logistic regression, simplified Apfel score and Koivuranta score.

Preoperative and intraoperative characteristics including gender, age, height, body weight, ASA status, duration of anaesthesia, type of surgery, history of motion sickness, history of previous PONV, smoking habit and use of postoperative opioid were counted as predictors (Tables 1 and 2). We used evaluations of χ^2 and information gain attributes to evaluate and select predictors to be used as input variables for ANN training. The information gain method is a measure motivated by information theory and it can be used to evaluate the worth of an attribute by measuring information gain with respect to the class. Information gain is the reduction of entropy (uncertainty) about the classification of a test class based on observation of a particular variable. It is the amount by which you reduce the uncertainty about the target class using or 'making a decision' about a particular variable. An irrelevant predictor was empirically defined as its information gain was lower than 5% of total information gains. Body weight, height, age and history of motion sickness were recognized as irrelevant predictors and were not used as input predictors because their values of information gain were below the selection threshold.

Neural computation was performed on an IBM compatible Pentium 4 computer running at 2.66 GHz. The ANN was constructed with neural network software (Statistica Neural Networks, Statsoft, Tulsa, OK, USA) using a training set of patients (n=656) who received general anaesthesia during

Table 1 Patient characteristics. Data are mean (SD) or numbers (percentage). PONV, postoperative nausea and vomiting; Note: Ophthalmology patients were excluded because of routine antiemetic prophylaxis

	Training set (n=656)	**Validation set (n=430)**
Number of cases PONV (%)	143 (21.8%)	104 (24.2%)
Age (yr)	45 (17)	38 (20)
Duration of anaesthesia (min)	100 (67)	115 (56)
Female	407 (62.0%)	274 (63.7%)
Male	249 (38.0%)	156 (36.3%)
Height (cm)	161.1 (8.6)	163.1 (9.7)
Body weight (kg)	62.6 (20.0)	59.2 (27.8)
ASA		
I	324 (49.4%)	183 (42.5%)
II	238 (36.3%)	180 (41.9%)
III	94 (14.3%)	67 (15.6%)
Non-smoker	446 (68.0%)	293 (68.1%)
History of motion sickness	206 (31.4%)	130 (31.9%)
History of previous PONV	49 (7.5%)	32 (7.4%)
Use of postoperative opioid	371 (56.6%)	250 (58.1%)
Type of surgery		
Orthopaedic	127 (19.4%)	85 (19.8%)
Ear, nose, throat (ENT)	74 (11.3%)	48 (11.2%)
General surgery	156 (23.8%)	110 (25.6%)
Gynaecological	162 (24.7%)	105 (24.4%)
Urological	37 (5.6%)	23 (5.3%)
Plastic surgery	66 (10.1%)	36 (8.4%)
Neurosurgery	19 (2.9%)	15 (3.5%)
Others	15 (2.3%)	8 (1.9%)

Table 2 Predictors calculated by χ^2 statistics and information gain with respect to PONV. N/A, not available. *These seven variables were used for prediction of PONV with the artificial neural network model

Predictors	**Pearson χ^2**	**Odds ratio**	**95% Confidence interval**	**Information gain**
Gender*	8.89	2.12	1.29–3.50	0.01561
Age	N/A	N/A	N/A	0
Height	N/A	N/A	N/A	0
Body weight	N/A	N/A	N/A	0
ASA status*	3.43	N/A	N/A	0.00543
Duration of anaesthesia*	N/A	N/A	N/A	0.02869
Type of surgery*	6.26	N/A	N/A	0.01149
History of motion sickness	0.01	0.974	0.601–1.577	0.00002
History of previous PONV*	12.56	3.52	1.69–7.33	0.01824
Smoking habit*	2.97	1.55	0.94–2.55	0.00514
Use of postoperative opioid*	14.25	2.52	1.55–4.11	0.02490

the first month of the study period. The predictors were entered into the ANN (as dichotomous yes/no input variables or numerical variables) as input variables. The presence of PONV was entered as a dichotomous yes/no output variable (Fig. 1). A multilayer perception (MLP) ANN with back propagation algorithm was used to train the predictive model. The MLP network consists of layers of virtual neurons in which neurons from contiguous layers are linked to each other by weighted connections. Each neuron in each layer receives as input a weighted sum of the outputs of all neurons in the previous layer. It then processes this input using a non-linear transfer function and transmits processed output to be used as input to neurons in the next layer. The configuration of the ANN was chosen by the software. The selected ANN consisted of one input layer, one hidden layer and one output layer. There were 21 neurons in input layer, 11 neurons in hidden layer and one neuron in output layer. During network training, a prediction was made and was correlated with the observed outcome. If the ANN predicted the outcome incorrectly, an error difference was calculated and the error was back-propagated through the network to adjust the connection weights to more closely match input and output data. A 10-fold cross-validation was used for evaluation. The training set was divided randomly into 10 subsets. Nine of the 10 subsets were used for training and the 10th subset was used for evaluation during training. The entire process was repeated nine additional times by rotating the subset that was used as the evaluation set. The mean square errors of each iteration were computed and averaged, and the ANN that had a mean square error closest to the average was selected.

After the training process finished, the network was validated with the remaining patients (n=430) from the original sample, who were not selected for training and whose outcome regarding PONV was unknown to the ANN.

The Naïve Bayesian approach is based on subjective probabilities. It has usually been used to build models for medical diagnosis. The predictors we used to calculate the probability of PONV were the same as those used with the ANN.

Logistic regression models the probability of some event occurring as a linear function of a set of predictor variables. The actual state of the dependent variable is determined by looking at the estimated probability. All of the predictors were chosen by fitting a logistic regression using a stepwise forward selection procedure (P<0.05 to enter).

Apfel and colleagues[16] identified four risk factors that predict PONV: female gender, prior history of PONV or motion sickness, non-smoking and the use of postoperative opioids. These risk factors had equal weights and were used to calculate the probability of PONV using a logistic regression model. The regression coefficients were calibrated with the training set.

Based on logistic regression analysis, Koivuranta identified five predictors for PONV, each with equal weight: female gender, previous PONV, duration of surgery over

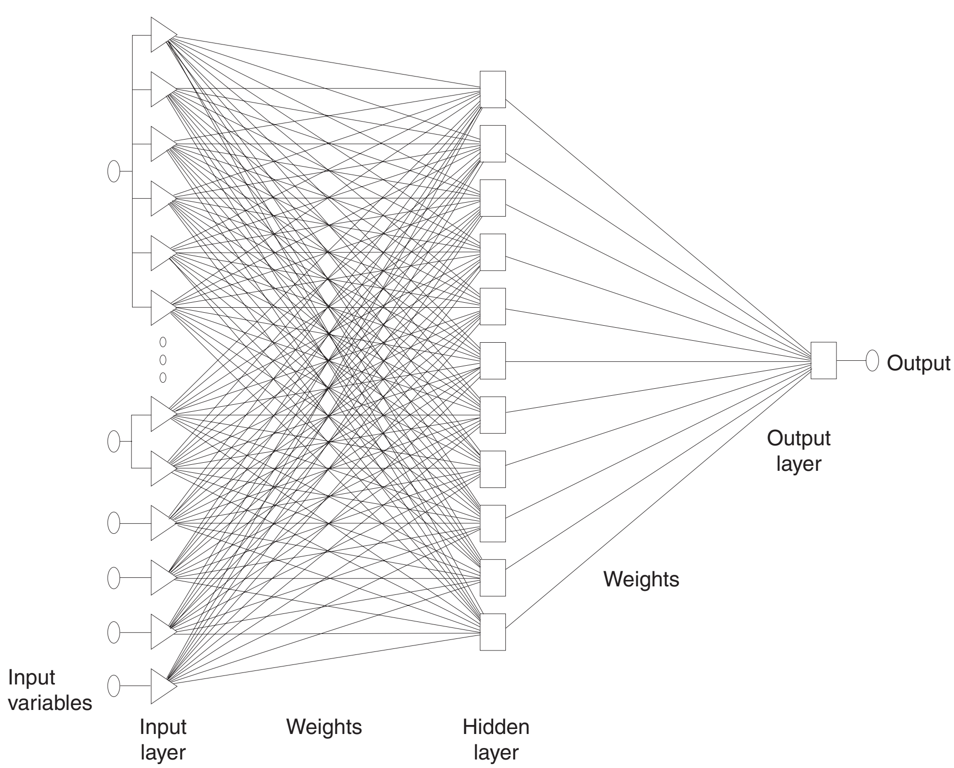

Fig 1 Illustration of the artificial neural network (ANN) model.

60 min, history of motion sickness and non-smoking.[8] Probability was calculated similarly to that for the simplified Apfel score.

Statistical analysis

Predictions were made with validation set using ANN, Naïve Bayesian classifier, logistic regression, Koivuranta score and simplified Apfel score. Accuracy (the number of correct predictions divided by total predictions), sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV) and likelihood ratios for positive and negative predictions of these models were calculated.

The discriminating power of these prediction models can be measured by receiver operating characteristic (ROC) curves. ROC analysis estimates a curve that describes the inherent trade-off between sensitivity and specificity of a prediction tool. Discriminatory power is measured by the area under the ROC curve (AUC). AUC is a particularly important metric for evaluating prediction tools because it is the average sensitivity over all possible specificities. AUC may range from 0 to 1, with area of 1.0 representing perfect discrimination and an area of 0.5 representing what is expected by chance alone.[17] In this study, a criterion (cut-off) value corresponding with the highest accuracy (minimal false negative and false positive results) was selected to report the sensitivity and specificity of each prediction tool.

Results

The incidence of PONV in the training group and validation group were 21.8% and 24.2%, respectively. The characteristics, surgical and anaesthesia data of the training group and validation group are shown in Table 1. In the training set, women had a 2-fold higher rate of PONV compared with men (odds ratio=2.12). Non-smokers (odds ratio=1.55), patients with a history of previous PONV (odds ratio=3.52) 3.52) and patients who received postoperative opioids (odds ratio=2.52) were also associated with a higher risk for PONV. Of the total of 1086 patients, 656 were selected for training the ANN and the rest (430 patients) were available for validation. The ANN had an overall accuracy of 83.3% in predicting PONV. The sensitivity was 77.9%, specificity was 85.0%, PPV was 62.3% and NPV was 92.3% (Table 3). The accuracy of the Naïve Bayesian classifier was 77.0%; the sensitivity was 18.3%, with a specificity of 95.7%, PPV of 57.6% and NPV of 78.6%. The accuracy of the logistic regression was 64.2%; the sensitivity was 62.5%, with a specificity of 64.7%, PPV of 36.1% and NPV of 84.4%. The accuracy of the Koivuranta score was 59.5%, with a sensitivity of 66.3%, specificity of 57.4%, PPV of 33.2% and NPV of 84.2%. The accuracy of the simplified Apfel score was 61.0%, with a sensitivity of 59.6%, specificity of 61.0%, PPV of 32.8% and NPV of 82.5%.

The ANN had the highest accuracy, sensitivity and PPV among all tested models. Table 3 shows the performance of ANN, Naïve Bayesian classifier, logistic regression, Koivuranta score and simplified Apfel score. The ANN had the best discriminating power with an AUC of 0.814. The AUC of the simplified Apfel score was 0.624, whereas the AUC of the Koivuranta score was 0.626, the AUC of the logistic regression was 0.669 and that of the Naïve Bayesian classifier was 0.570.

The performance of these models was compared with each other by using pairwise comparison of ROC curves (Table 4). We found that there were no statistically significant differences among the simplified Apfel score, Koivuranta score and logistic regression. The ANN model (AUC=0.814) performed significantly better than the other models ($P<0.05$).

Discussion

Differing from the studies of Eberhart and Traeger, our study did not restrict the anaesthetic techniques and agents, and the predictors were selected by information gain instead of odds ratio. In this study, the ANN was superior to the simplified Apfel score, Koivuranta score, logistic regression and Naïve Bayesian classifier in predicting PONV. The ability of the ANN to process more information in the context of multidimensionality of complex data to a large extent explains the superiority of ANN as a predicting model. The screen of risk factors in predicting PONV was important in

Table 3 Comparison of predictive performance of artificial neuronal network (ANN), Naïve Bayesian classifier, Koivuranta score and simplified Apfel score. LR+ and LR−, positive and negative likelihood ratios; NPV, negative predictive value; PPV, positive predictive value; ROC, receiver operating curve

	ANN	Logistic regression	Naïve Bayesian classifier	Koivuranta score	Simplified Apfel score
Accuracy (%)	83.3	64.2	77.0	59.5	61.0
Sensitivity (%)	77.9 (68.7–85.4)	62.5 (52.5–71.8)	18.3 (11.4–27.1)	66.3 (56.4–75.3)	59.6 (49.5–69.1)
Specificity (%)	85 (80.6–88.7)	64.7 (59.3–69.9)	95.7 (92.9–97.6)	57.4 (51.8–62.8)	61.0 (55.5–66.4)
PPV (%)	62.3	36.1	57.6	33.2	32.8
NPV (%)	92.3	84.4	78.6	84.2	82.5
LR+	5.18	1.77	4.25	1.56	1.53
LR−	0.26	0.58	0.85	0.59	0.66
Area under ROC curve	0.814 (0.77–0.85)	0.669 (0.62–0.71)	0.570 (0.52–0.62)	0.626 (0.58–0.67)	0.624 (0.58–0.67)
Standard error	0.027	0.032	0.033	0.033	0.033

Table 4 Pairwise comparison of ROC curves. ROC, receiver operating curve; ANN, artificial neural network; AUC, area under the ROC curve

Pairwise comparison	Difference between AUC	Standard error	95% Confidence interval	*P*-value
ANN *vs* Naïve Bayesian classifier	0.244	0.039	0.167–0.322	<0.05
ANN *vs* simplified Apfel score	0.190	0.041	0.111–0.270	<0.05
ANN *vs* Koivuranta score	0.188	0.041	0.109–0.268	<0.05
ANN *vs* logistic regression	0.145	0.038	0.070–0.220	<0.05
Simplified Apfel score *vs* Naïve Bayesian classifier	0.054	0.037	−0.019–0.127	0.146
Simplified Apfel score *vs* Koivuranta score	0.002	0.020	−0.036–0.040	0.923
Simplified Apfel score *vs* logistic regression	0.045	0.025	−0.004–0.094	0.071
Naïve Bayesian classifier *vs* Koivuranta score	0.056	0.038	−0.017–0.130	0.135
Naïve Bayesian classifier *vs* logistic regression	0.100	0.035	0.030–0.169	<0.05
Koivuranta score *vs* logistic regression	0.043	0.034	−0.022–0.109	0.196

building the ANN model. Using too many predicting variables or using predictors that have little influence in PONV would contribute to a lower accuracy with the ANN. We collected patient-related and anaesthesia-related predictors that had been used in other studies to build predicting models.[11] To increase the accuracy and efficiency of our ANN, we selected predictors as input variables by calculating the value of information gain and using χ^2 statistics. Of the 11 predictors, 7 variables—gender, type of surgery, ASA status, duration of anaesthesia, smoking habit, history of previous PONV and use of postoperative opioid were used to train the ANN model. The ANN is a dynamic model; with each new patient the model back-propagates and checks data with an error-minimization function, re-adjusting hidden weights to improve predictive accuracy (Fig. 1). Thus, as data for more and more patients were entered into the model, self-learning and error correction by back-propagation and in turn predictive accuracy, improved progressively. The major advantage of using an ANN in predicting PONV is that neural networks train themselves without much human intervention.

ANNs are computer models composed of parallel, non-linear computational neurons arranged in highly interconnected layers. They can define relationships among input data that are not apparent when using traditional statistical approaches, and they can use these relationships to improve accuracy. Hence, neural nets have substantial power to identify patterns in complex datasets. Although the networks are conceptually attractive, ANNs are not analysed easily based on risks attributable to specific clinical characteristics or statistical significance. This is because a neural network rests on its internal representation of weights and functions to process data instead of straightforward equations such as a regression model. The ANN model is an algorithm that can be used to perform non-linear statistical modelling and provides a new alternative to logistic regression, the most commonly used method for developing predictive models for dichotomous outcomes in the field of medicine. The advantages of the ANN include requiring less formal statistical training, ability to identify complex non-linear relationships between dependent and independent variables, ability to detect all possible interactions between predictor variables, and availability of multiple training algorithms. Disadvantages include its 'black box approach', greater computational burden, proneness to over-fitting and the empirical nature of model development.[18]

It is still unclear to what extent the weights attributed to risk factors are based on multivariable analysis in predicting PONV. Several risk factors show a tendency to be more frequent in one group than in the other, and this may lead to a significantly different risk when assessed by a scoring model. Therefore, unless significantly stronger predictors are identified that appear to be applicable to most patients, it seems unlikely that any scoring predictive models will lead to significant better prediction of PONV. The ANN can be trained in different institutions to develop different models that fit the characteristics and patterns of their own patients. The predictions made by these ANN models would be more precise than that made by universally using a score model.

In summary, routine medications to prevent PONV are not recommended for several reasons including economic reasons, potential side-effects of antiemetic drugs and lack of increased patient satisfaction.[19 20] Thus, identifying high-risk patients before preventive pharmacological interventions are prescribed becomes very important. A good predictive model for PONV needs high sensitivity, specificity, PPV and NPV. The ANN had the best accuracy, sensitivity, PPV and NPV of models evaluated in this study. Computer-based medical decision support systems have recently been studied and used clinically for medical diagnosis and improvement of patient care.[21] The ANN used in this study can be easily used with any standard desktop computer. Our study did not have an external validation; hence we were not able to assure that ANN model developed in a particular institution performed as well at another institution. Nevertheless, the ANN can be easily developed in any institution for local use. Therefore, the ANN appears to be a very suitable model for clinicians to use in putting rational and cost-effective antiemetic treatments into practice. Prospective clinical studies will be needed to evaluate the use of such models in improving cost-effective prophylactic intervention for PONV according to the predictions of the ANN.

References

1 Watcha MF, White PF. Postoperative nausea and vomiting. Its etiology, treatment and prevention. *Anesthesiology* 1992; **77**: 162–84

2 Cohen MM, Duncan PG, Deboer DP, Tweed WA. The postoperative interview: assessing risk factors for nausea and vomiting. *Anesth Analg* 1994; **78**: 7–16

3 Quinn AC, Brown JH, Wallace PG, Asbury AJ. Studies in postoperative sequelae. Nausea and vomiting—still a problem. *Anaesthesia* 1994; **49**: 62–5

4 Lerman J. Surgical and patient factors involved in postoperative nausea and vomiting. *Br J Anaesth* 1992; **69**: 24–32

5 Rabey PG, Smith G. Anaesthetic factors contributing to postoperative nausea and vomiting. *Br J Anaesth* 1992; **69**: 40–5

6 Kapur PA. The big 'little problem'. *Anesth Analg* 1991; **73**: 243–5

7 Wilder-Smith OH, Martin NC, Morabia A. Postoperative nausea and vomiting: a comparative survey of the attitudes, perceptions, and practice of Swiss anesthesiologists and surgeons. *Anesth Analg* 1997; **84**: 826–31

8 Koivurantu M, Läärä E, Snare L, Alahuta S. A survey of postoperative nausea and vomiting. *Anaesthesia* 1997; **52**: 443–9

9 Gan TJ, Meyer T, Apfel CC, *et al*. Consensus guideline for the management of postoperative nausea and vomiting. *Anesth Analg* 2003; **97**: 62–71

10 Apfel CC, Kranke P, Eberhart LHJ, Roos A, Roewer N. Comparison of predictive models for postoperative nausea and vomiting. *Br J Anaesth* 2002; **88**: 234–40

11 Thomas R, Jones NA, Strike P. The value of risk scores for predicting postoperative nausea and vomiting when used to compare patient groups in a randomised controlled trial. *Anaesthesia* 2002; **57**: 1119–28

12 Eberhart LHJ, Högel J, Seeling W, Staack AM, Geldner G, Georgieff M. Evaluation of three risk scores to predict postoperative nausea and vomiting. *Acta Anaesthesiol Scand* 2000; **44**: 480–8

13 Cross SS, Harrison RF, Kennedy RL. Introduction to neural networks. *Lancet* 1995; **346**: 1075–9

14 Eberhart L, Traeger M, Nawroth A, *et al*. Using an artificial neuronal network to predict postoperative nausea and vomiting. A comparison with two conventional risk scores based on logistic regression analysis. *Anesthesiology* 2001; **95**: A1104

15 Traeger M, Eberhart A, Geldner G, *et al*. Prediction of postoperative nausea and vomiting using an artificial neural network. *Anaesthesist* 2003; **52**: 1132–8

16 Apfel CC, Läärä E, Koivuranta M, Greim CA, Roewer N. A simplified risk score for predicting postoperative nausea and vomiting. *Anesthesiology* 1999; **91**: 693–700

17 Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* 1982; **143**: 29–36

18 Tu JV. Advantages and disadvantages of using artificial neural networks versus logistic regression for predicting medical outcomes. *J Clin Epidemiol* 1996; **49**: 1225–31

19 Sung YF. Risks and benefits of drugs used in the management of postoperative nausea and vomiting. *Drug Saf* 1996; **14**: 181–97

20 Scuderi PE, James RL, Harris L, Mims GR. Antiemetic prophylaxis does not improve outcomes after outpatient surgery when compared to symptomatic treatment. *Anesthesiology* 1999; **90**: 360–71

21 Johnston ME, Langton KB, Haynes B, Mathieu A. Effects of computer-based clinical decision support systems on clinician performance and patient outcome: a critical appraisal of research. *Ann Intern Med* 1994; **120**: 135–42