The Journal of International Medical Research
2009; 37: 1680 – 1691 [first published online as 37(6) 3]

Machine Learning Methods Applied to Pharmacokinetic Modelling of Remifentanil in Healthy Volunteers: a Multi-method Comparison

MR Poynton[1,*], BM Choi[2,*], YM Kim[3], IS Park[4], GJ Noh[5], SO Hong[6], YK Boo[7] and SH Kang[8]

[1]Informatics Program, University of Utah College of Nursing and Department of Biomedical Informatics, University of Utah School of Medicine, Salt Lake City, Utah, USA; [2]Department of Anaesthesiology and Pain Medicine, National Medical Centre, Seoul, Republic of Korea; [3]Korea Health Industry Development Institute, Seoul, Republic of Korea; [4]National Health Insurance Corporation, Seoul, Korea; [5]Department of Clinical Pharmacology and Therapeutics, Department of Anaesthesiology and Pain Medicine, Asan Medical Centre, University of Ulsan College of Medicine, Seoul, Republic of Korea; [6]Department of Chronic Disease Surveillance, Korea Centres for Disease Control and Prevention, Ministry for Health, Welfare and Family Affairs, Seoul, Republic of Korea; [7]College of Health Industry, Eulji University, Seongnam, Republic of Korea; [8]School of Health Administration, Inje University, Kimhae, Republic of Korea

This study compared the blood concentrations of remifentanil obtained in a previous clinical investigation with the predicted remifentanil concentrations produced by different pharmacokinetic models: a non-linear mixed effects model created by the software NONMEM®; an artificial neural network (ANN) model; a support vector machine (SVM) model; and multi-method ensembles. The ensemble created from the mean of the ANN and the non-linear mixed effects model predictions achieved the smallest error and the highest correlation coefficient. The SVM model produced the highest error and the lowest correlation coefficient. Paired *t*-tests indicated that there was insufficient evidence that the predicted values of the ANN, SVM and two multi-method ensembles differed from the actual measured values at $\alpha = 0.05$. The ensemble method combining the ANN and non-linear mixed effects model predictions outperformed either method alone. These results indicated a potential advantage of ensembles in improving the accuracy and reducing the variance of pharmacokinetic models.

KEY WORDS: Artificial intelligence; Neural networks (computer); Pharmacokinetics; Remifentanil

Introduction

In pharmacokinetic modelling, artificial neural networks (ANNs) have demonstrated equivalent or greater accuracy than the dominant statistical modelling method, non-linear mixed effects modelling, as

*MR Poynton and BM Choi contributed equally to this article.

implemented by the popular non-linear mixed effects model software package, NONMEM®.[1,2] In a recent study by Kang *et al.*,[3] ANN models produced more accurate predictions of blood concentrations of remifentanil than a non-linear mixed effects model. This earlier work also demonstrated an important use for accurate predictions of individual blood concentrations of remifentanil: when paired with an electroencephalographic parameter, these predictions could be used to create an improved pharmacodynamic model of the effect of remifentanil on the central nervous system.[3] One of the recently emerging powerful modelling tools for non-linear relationships is support vector machines (SVMs), which are useful analytical tools for population pharmacokinetic modelling.[4]

Studies comparing the use of multiple methods, such as SVMs and ensemble methods, on the same pharmacokinetic data are scarce. The comparative accuracy of SVMs and ensemble techniques for pharmacokinetic modelling relative to non-linear mixed effects models and ANNs has not been established. These methods are potentially useful for the purpose of pharmacokinetic modelling if their accuracy exceeds or is equivalent to that of non-linear mixed effects models. As an extension of the work by Kang *et al.*,[3] the present study aimed to compare the accuracy of non-linear mixed effects models, ANNs and SVMs, as well as multi-method ensembles, in predicting blood concentrations of the opioid anaesthetic remifentanil.

Subjects and methods

SUBJECTS

The present retrospective analysis included data collected from healthy volunteers during a previous study, where blood concentrations of remifentanil were obtained.[5] After obtaining the approval from the Asan Medical Centre Institutional Review Board (IRB) and written informed consent from volunteers, those volunteers with no medical problems or abnormal laboratory test results were enrolled in this previous clinical investigation, which was conducted at the Asan Medical Centre from March 2005 to November 2005.[5] The Asan Medical Centre IRB agreed to the induction of muscle paralysis in volunteers given remifentanil in doses that may have been too low to guarantee unconsciousness; all volunteers were clearly informed and agreed that they might be awake and unable to move or breathe spontaneously during the course of the study. Volunteers were divided into three groups according to age: young (19 – 40 years old), middle-aged (41 – 64 years old) or elderly (≥ 65 years old).

INFUSION RATES

Young volunteers were randomly assigned to receive remifentanil at a constant rate of infusion of 1, 2, 3, 4, 5, 6, 7 or 8 μg/kg per min for 20 min. Middle-aged or elderly volunteers received remifentanil at the rate of 3 mg/kg per min until the real-time 95% spectral edge frequency (SEF_{95}) did not show further change after reaching maximal suppression. Because high-dose remifentanil may cause severe muscle rigidity, interfering with spontaneous breathing as well as assisted ventilation by anaesthesiologists, a neuromuscular blocking agent was administered to all volunteers that prohibited subjects from breathing spontaneously and necessitated manual ventilation by anaesthesiologists. Arterial blood samples (3 ml) were taken at the following intervals: every 30 s during the first 5 min, every 1 min during the second 5 min and every 2 min during the third 10 min after the beginning of remifentanil infusion; every 30 s during the first 5 min, every 1 min

during the second 5 min and every 2 min during the third 10 min after the termination of remifentanil infusion; and at 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 120, 150, 180, 210 and 240 min after the end of remifentanil infusion.

DATA SETS

Each instance in the data sets consisted of values for the subject's pharmacokinetic parameters and a measured concentration of remifentanil. Because of variability in the number of instances and the time point for blood concentration measurement, each instance was treated independently. Data preparation and analyses were conducted using NONMEM® VI level 2 (GloboMax, Ellicott City, MD, USA), SAS version 4.1 (SAS Institute, Cary, NC, USA) and STATISTICA software, version 7.1 (StatSoft, Tulsa, OK, USA). The predictor attributes used were time elapsed since initial infusion (min), total amount of agent infused (mg), rate of infusion (mg/min), body weight (kg), height (cm), body surface area (m^2), lean body mass (kg), infusion (during infusion or post-infusion) and gender.

DATA PARTITION

The performance of each method was assessed by five-fold cross-validation with random subsampling, in which 80% and 20% of the data were used as a training and a testing set, respectively. For five-fold cross-validation, the data set was split into five mutually exclusive subsets (the folds) of approximately equal size – D1, D2, D3, D4 and D5 – using random subsampling in SAS version 4.1. The homogeneity of the five folds was examined using an analysis of variance procedure and no statistically significant differences were found. Since the model was developed using 80% of the data and tested in the remaining 20%, this meant that in the course of the five experiments each instance was used four times for training and once for testing. This technique maximized the amount of data available for training (using every available instance) and also enabled the calculation of performance estimates (estimates of accuracy) using every instance available in the limited data set. Performance estimates for each method were calculated from the cumulative results for all five folds.

NON-LINEAR MIXED EFFECTS MODEL

The non-linear mixed effects model was fitted using the first-order estimation method. A full variance–covariance matrix was estimated for the different distributions of η. The models were evaluated using statistical and graphical methods. The minimal value of the objective function (OFV, equal to minus twice the log likelihood) was used as the goodness-of-fit characteristic to discriminate between hierarchical models using the log likelihood ratio test.[6] A *P*-value of 0.05, representing a decrease in OFV of 3.84 points, was considered to be statistically significant (χ^2 distribution, one degree of freedom). Individual Bayesian pharmacokinetic parameters were obtained using the POSTHOC option.[6] The S-plus® (MathSoft, Seattle, WA, USA) based model-building aid Xpose 3.1 was used for graphical model diagnosis.[7] In this study, the concentration–time data were best described by a three compartment pharmacokinetic model which was fitted using the ADVAN11 TRANS1 subroutine of NONMEM®. The covariates analysed were age, gender, weight, height, body surface area and lean body mass. Scatter-plots of covariates against initial parameter estimates were examined to identify those factors that may have a potential influence in the model. Each

covariate was then added sequentially to the initial model. Covariates significantly influencing the model were added in a forward selection. Backward elimination was then performed by fixing the coefficient of each covariate in turn to zero. Normality of the distribution of η was verified by testing the null hypothesis that the distribution of the data does not differ from that of a normal distribution.[8] Non-parametric bootstrap analysis was performed as an internal model validation, using the software package Wings for NONMEM® (Holford NH, version 614, July 2007, Auckland, New Zealand).[9]

NONMEM® produces predicted and individually predicted values (*post hoc* Bayesian estimates). For the individually predicted values, random inter-individual variability, η_i, is realized, which should improve the fitting performance of a model. For the predicted values, however, η_i is assumed to be zero with variance ω^2. In the present study predicted concentrations of remifentanil were used. The inter-individual random variability on each of the model parameters were modelled assuming a log-normal distribution:

$$P_i = \theta_{TV} e^{\eta_i} \quad (1)$$

where P_i is the parameter of an individual, θ_{TV} is a fixed effect parameter in a typical individual (*TV* denotes a typical value), e is the base of the natural logarithm and η_i is a random variable that describes the inter-individual variability of the parameter across the population. The residual random variability was modelled with an additive plus proportional error model. The median-weighted residual and median absolute weighted residual for the final pharmacokinetic models were calculated to examine the quality of prediction for the population. The weighted residual was calculated as the (measured – predicted)/predicted.[10]

ARTIFICIAL NEURAL NETWORK

The ANN employed a multilayer perception architecture with one hidden layer consisting of 12 processing units. Inputs were normalized with the mean = 0 and SD = 1 in order to prevent dominance of attributes with broad ranges of values and/or extreme values. The Levenberg–Marquardt method of weight adjustment was used to train or optimize parameters for the ANN model. The formula for weight adjustment using the Levenberg–Marquardt method is as follows:

$$w_{l+1} = w_l - (J^T J + \mu I)^{-1} J^T e \quad (2)$$

where w_{l+1} and w_l are the network weights, J is the Jacobian matrix that contains first derivatives of the network output with respect to each weight estimate, J^T is the transpose matrix of J, e is the vector of network residuals, μ is the learning parameter and I is the identity matrix.[11] The Levenberg–Marquardt method was chosen for this task because it makes very small weight adjustments as it minimizes error, finding the most correct or optimal parameters. It also works well in situations where parameter estimates are highly correlated, a possibility in this data set with multiple attributes/inputs describing body composition.[11]

Artificial neural network models are prone to over-fitting when model parameters fit the training data so closely that the model converges on idiosyncrasies of the data set or random variance rather than generalizable phenomena. In order to prevent over-fit, performance was estimated using five-fold cross-validation. The five models, each built using a distinct testing set, were structurally similar.

SUPPORT VECTOR MACHINE

The SVM model used was a type I regression SVM.[12] SVM parameters were selected using a grid search and five-fold cross-validation. The type I or C-SVM minimizes the following

error function:

$$\tfrac{1}{2}w^T w + C\sum_{i=1}^{N}(\xi_i + \xi_i^*) \quad (3)$$

subject to the constraints:

$(w^T\phi(x_i) + b) - y_i \le \varepsilon + \xi_i,\ \ \xi_i, \xi_i^* \ge 0$, and

$i = 1,\ldots\ldots,N$ (4)

where C is the capacity constant, w is the vector of coefficients, b is a constant, y_i represents targets for all the training data, ε is the insensitivity parameter, ξ_i are parameters for handling inputs, ϕ is the kernel function, x_i represents the independent variables, and the index i labels the number (N) of training cases.

Linear, polynomial, radial basis function (RBF) and sigmoid kernel functions were implemented. Error served as the objective criteria for selecting kernel function (ϕ). The parameters C, γ or kernel parameters and the insensitivity parameter ε in equation 4 were selected using a grid search and five-fold cross-validation. Although each of five experiments yielded a slightly different SVM, an SVM using the RBF kernel $\gamma = 0.1$, $C = 10$ and $\varepsilon = 0.1$ produced minimal error.

ENSEMBLE MODELS

Two simple ensembles were created using averaging of multiple methods. Ensemble I was the mean of the predicted values from all three models (non-linear mixed effects model, ANN, SVM). Ensemble II was the mean of the non-linear mixed effects model and ANN predicted values.

COMPARISON OF ACCURACY

Accuracy estimates for all models were calculated using five-fold cross-validation, as previously described; these were calculated using the cumulative test results from all folds. Accuracy was compared using three measures: mean absolute error (MAE), mean squared error (MSE) and the correlation coefficient for predicted versus actual blood concentration values. For each method, a scatter-plot was used to visualize the distribution of predicted versus actual blood concentrations. Histograms and pairwise residual plots were used to visualize and compare the distributions of predicted values. Paired *t*-tests were used to test for a difference in predictive accuracy among the five methods – non-linear mixed effects model, ANN, SVM, Ensemble I and Ensemble II – and the measured values.

Results

A total of 30 volunteers were enrolled in a previous study[5] from whom remifentanil pharmacokinetic data were obtained.

Volunteer characteristics are given in Table 1. Body weight, height, body surface area and lean body mass were not significantly different among the three groups. It is unlikely, therefore, that middle-aged or elderly volunteers resembled each other more than the younger population.

A total of 1340 instances were recorded for the 30 subjects. The predictor attributes are given in Table 2. The blood concentrations of remifentanil over time in four of the volunteers are shown Fig. 1.

Goodness-of-fit statistics for the different pharmacokinetic models are presented in Table 3, and scatter-plots of predicted versus measured values are given in Fig. 2. Among the five models, Ensemble II (the mean of the non-linear mixed effects model and ANN predictions) achieved the smallest error and the highest correlation coefficient. The SVM achieved the highest error measurements and lowest correlation coefficient. The scatter-plots visually support these results (Fig. 2).

Paired *t*-tests indicated that there was insufficient evidence to conclude that the predictions of the SVM, ANN, Ensemble I or Ensemble II models differed from the measured values (Table 4). There was,

TABLE 1:
Characteristics of study volunteers divided according to age into young (19 – 40 years old), middle-aged (41 – 64 years old) and elderly (≥ 65 years old) groups

Characteristic	Young (*n* = 10)	Middle-aged (*n* = 10)	Elderly (*n* = 10)
Sex			
Male	5	5	5
Female	5	5	5
Weight (kg)	62.2 ± 11.7	59.5 ± 12.7	57.8 ± 8.4
Height (cm)	166.0 ± 9.4	163.5 ± 10.0	158.3 ± 8.5
Age (years)	27.0 ± 6.0	45.6 ± 4.5	71.2 ± 4.4
Body surface area (m^2)	1.69 ± 0.19	1.64 ± 0.22	1.58 ± 0.14
Lean body mass (kg)	48.4 ± 9.5	46.6 ± 10.2	44.2 ± 6.8
Infusion rate (μg/kg per min)	4.1 ± 2.3	3.0[a]	3.0[a]
Infusion duration (min)	20.0 ± 0.2	15.6 ± 4.2	15.4 ± 3.9
Total amount of remifentanil (mg)	5.1 ± 2.9	2.8 ± 1.0	2.6 ± 0.7

[a]Fixed rate infusion of remifentanil.
Data are number of subjects or mean ± SD.

however, evidence that the non-linear mixed effects model predictions differed from the measured values: the paired *t*-test for non-linear mixed effects predicted values versus measured values reached statistical significance at $\alpha = 0.05$.

Discussion

Population pharmacokinetics and pharmacodynamics can be defined as the study of the variability in drug concentration or effect between individuals when standard dosage regimens are administered. Population analysis measures this variability within the population and accounts for it in terms of patient-specific covariates, including age, weight, disease state and concurrent therapy. Pharmacokinetic modelling invites the consideration and comparison of prospective alternative modelling methods for several reasons, in particular the challenging characteristics of

TABLE 2:
Predictor attributes used for the data sets

Input	Description	Type	*r*
Time	Time elapsed since initial infusion (min)	Continuous	–0.44
AMT	Total amount of agent infused (mg)	Continuous	0.37
Rate	Rate of infusion (mg/min)	Continuous	0.38
BWT	Body weight (kg)	Continuous	0.17
HT	Height (cm)	Continuous	0.13
BSA	Body surface area (m^2)	Continuous	0.17
LBM	Lean body mass (kg)	Continuous	0.16
InfYes	Infusion (during infusion or post-infusion)	Binary	–
Sex	Gender of the subject (male or female)	Binary	–

r, Pearson's correlation coefficient describing correlation between input and measured concentration of remifentanil; all correlations were significant at $P < 0.05$.

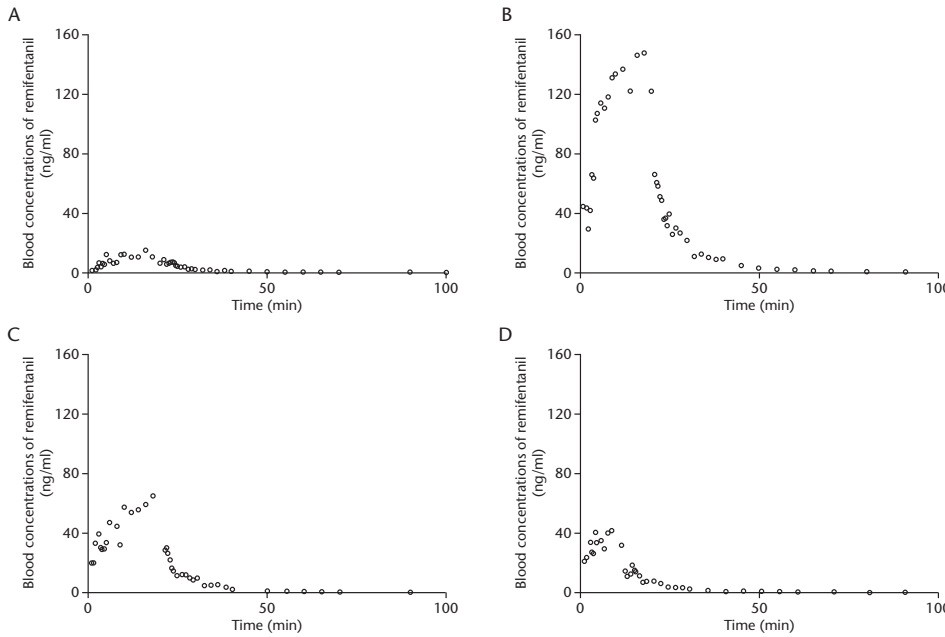

FIGURE 1: Blood concentrations of remifentanil over time: (A) volunteer ID 2 (24 years old, 63 kg, 1 mg/kg per min remifentanil for 20 min); (B) volunteer ID 4 (26 years old, 71 kg, 7 mg/kg per min remifentanil for 20 min); (C) volunteer ID 8 (20 years old, 63 kg, 3 mg/kg per min remifentanil for 20 min), (D) volunteer ID 13 (56 years old, 59.8 kg, 3 mg/kg per min remifentanil for 10.8 min)

pharmacokinetic data. These data sets consist of repeated measures in time series, they are typically imbalanced due to unequal numbers of measurements among subjects and the measurements are not taken at common time points. These considerations are normally addressed with non-linear mixed effects models.

The non-linear mixed effects method analyses the data of many individuals at once, but takes the inter-individual random effects structure into account. It is generally assumed that the underlying structure of the pharmacokinetic or pharmacodynamic

TABLE 3:
Goodness-of-fit statistics for the different pharmacokinetic models

Pharmacokinetic model	MSE	MAE	r^2
SVM	181.20	9.72	0.82
NONMEM	95.77	5.43	0.91
ANN	57.12	4.10	0.95
Ensemble I (SVM, NONMEM, ANN)	70.80	5.34	0.93
Ensemble II (NONMEM, ANN)	55.17	4.17	0.95

MSE, mean squared error; MAE, mean absolute error; r^2, correlation coefficient; SVM, support vector machine; NONMEM, non-linear mixed effects model created using the software NONMEM®; ANN, artificial neural network.

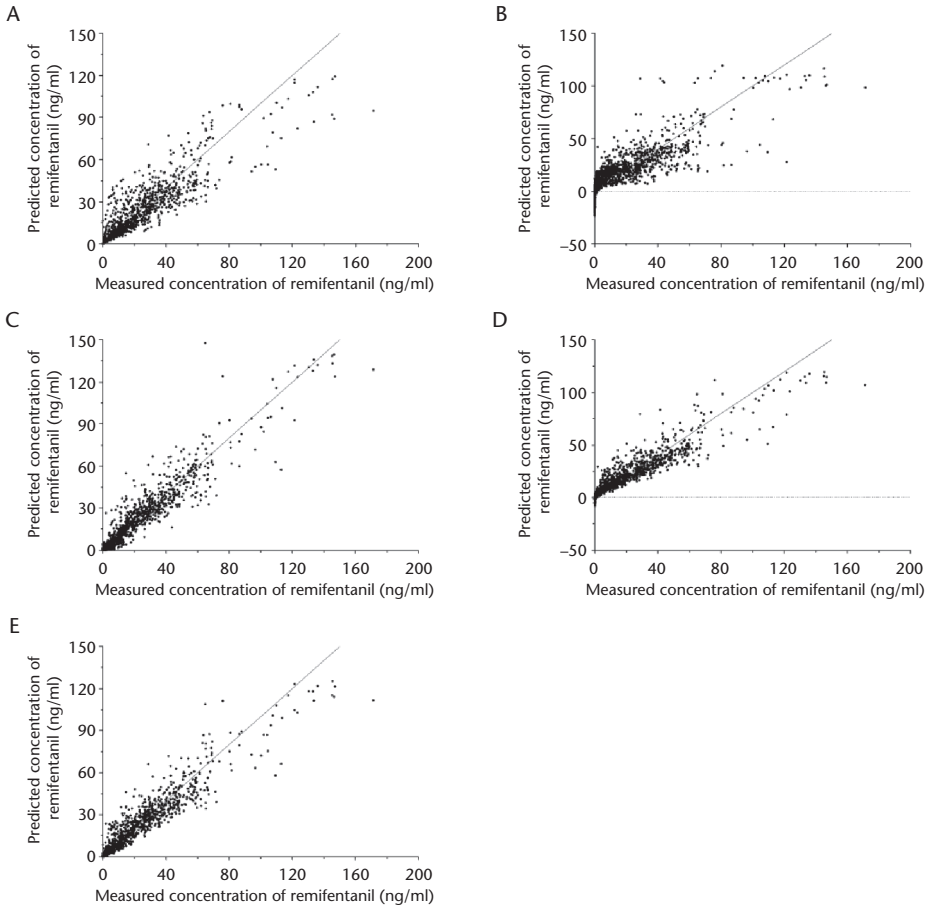

FIGURE 2: Goodness-of-fit plots of predicted versus measured blood concentrations of remifentanil for all the models: (A) non-linear mixed effects model created using the software NONMEM®; (B) support vector machine model (SVM); (C) artificial neural network model (ANN); (D) Ensemble I model (SVM, NONMEM, ANN); (E) Ensemble II model (NONMEM, ANN)

model is similar for each individual, whereas the actual value of the parameter vector differs from individual to individual. The components necessary to describe a population pharmacokinetic or pharmacodynamic model include a model describing the expected response (drug concentration or effect) in a specific individual, a model describing the relationship between patient characteristics and the pharmacokinetic or pharmacodynamic model parameters, a model for the inter-individual random variability in the patient population and, finally, a model for the random residual variability in the data.[13]

Non-linear mixed effects models may be successfully fitted to pharmacokinetic data because they accommodate both repeated measures and unbalanced data, and because they are capable of fitting models when there are only sparse data.[14] They also accommodate the non-linear relationship

TABLE 4:
Paired *t*-tests of pharmacokinetic model predictions versus measured values of remifentanil

Pharmacokinetic model	Mean	SD	*t*-value
SVM	18.99	21.38	–1.10
NONMEM	17.95	20.74	2.39[a]
ANN	18.46	22.85	0.60
Ensemble I (SVM, NONMEM, ANN)	18.47	20.72	0.52
Ensemble II (NONMEM, ANN)	18.21	21.33	1.88

[a]$P < 0.05$ for predictions versus measured values of remifentanil.
SVM, support vector machine; NONMEM, non-linear mixed effects model created using the software NONMEM®; ANN, artificial neural network.

between time and blood concentration of pharmacotherapeutic agents. A popular software tool optimized to fit non-linear mixed effects models to pharmacokinetic data is NONMEM®. Use of NONMEM® requires a thorough *a priori* understanding of the method, the software and the pharmaceutical agent for proper model construction. It is both time and labour intensive. The use of intelligent techniques – methods that automatically build models and adjust parameters – can substantially speed the modelling process and may offer improvements in predictive accuracy. Multimethod comparisons illustrating the relative strengths of the available methods are, however, scarce.

One alternative method for pharmacokinetic modelling – ANNs – have proven more accurate than non-linear mixed effects models in several studies.[1,2,15] In 1995, Brier *et al.*[15] published a study describing prediction of peak and trough blood concentrations of gentamicin in which an ANN and a non-linear mixed effects model constructed using NONMEM® software demonstrated equivalent accuracy. A study in 2000 by Tolle *et al.*[2] compared back-propagation neural network and non-linear mixed effects model predictions of blood concentrations of tobramycin in a paediatric population. Again, the ANN and the non-linear mixed effects model created using NONMEM® software demonstrated equivalent performance. ANNs, however, require adjustment of many parameters and are prone to over-fitting, when they fit a model to the training data so closely that they fit noise and error as well as true pattern. When this occurs, the model does not generalize well.

Support vector machines, a machine learning method developed by Vapnik *et al.*,[16] rival ANNs in their ability to approximate complex, non-linear decision boundaries. Several studies have shown equivalent accuracy of ANN and SVM models.[17,18] Some studies indicate, however, that SVMs are more accurate than ANNs.[19,20] This seems counterintuitive given the ability of ANNs to approximate any existing solution.[21] As previously mentioned, the flexibility of ANN architectures can lead to rote learning of a training data set or over-training. In these circumstances, the ANN does not generalize well to new data; when externally validated, accuracy deteriorates. In addition, ANNs require extensive parameterization and predictive accuracy may be suboptimal due to the selection of inappropriate parameters.

Support vector machines require the

selection of very few training parameters, limiting the opportunity for human error. For smaller, less dimensional data sets, it is computationally feasible to conduct a grid search for optimal parameters using cross-validation. Clinical data, especially data stored in a very granular form, are often sparsely populated. Because SVMs do not seek to minimize generalization error but, rather, they bound the limits of generalization error, multivariate modelling of sparse data is not problematic.[16] SVMs are capable of approximating complex, non-linear decision boundaries but are less prone to over-training than ANNs, again because they bound the upper limit of generalization error and do not seek to minimize error itself. Though it is possible to over-train a SVM, it is less likely and, therefore, a SVM may perform with greater accuracy than an ANN when deployed on new data.

Ensemble methods – methods that combine the predictions of multiple models – are yet another approach to producing numerical predictions. There are many ways to combine the outputs of multiple methods and models to produce a prediction. Models can be built in a modular fashion to perform tasks that can be used in combination to produce a prediction or classification. Alternatively, multiple models can be built for the same task, and their outputs combined to produce a decision.[22] One very simple and effective approach is that of averaging model outputs.[23]

In the present multi-method comparison, an ensemble model (Ensemble II) combining the non-linear mixed effects model and an ANN achieved the highest accuracy as measured by MSE, MAE and the correlation coefficient. A paired *t*-test failed to indicate that the predictions of the ANN and Ensemble II models differed, indicating there was no particular advantage in using an ensemble that incorporated the non-linear mixed effects predictions rather than the ANN predictions alone. The scatter-plots and residual plots showed a visual reduction in outliers – i.e. a reduction in variance. This would be desirable when a pharmacokinetic model is being used to predict blood concentrations for individuals, as it would be advantageous to reduce the size of residuals in order to avoid clinically important disparities between real and predicted concentrations. In particular, ANN excessively under-predicted the lower range of measured concentrations of remifentanil (especially < 0.4 ng/ml). In this range, there were no significant differences in predictive performance between the ANN and non-linear mixed effects models. The predicted concentration at 0.08 ng/ml of measured concentration of remifentanil for one point in the non-linear mixed effects model was 16.651 ng/ml. If this was considered an outlier and excluded from predictive performance analysis, there would be high MSE and MAE values in the ANN pharmacokinetic model compared with the non-linear mixed effects model ($P = 0.019$, two sample *t*-test). This point was, however, included in all analyses because it was not measurement error, which may explain why there were no significant differences between the predictions of the ANN and Ensemble II models.

The availability of measured blood concentrations as a gold standard was an important and uncommon reference point for evaluating the accuracy of the different modelling methods. The statistical approach chosen for this study evaluated the accuracy of multiple methods against the two best available reference points: the measured blood concentrations of remifentanil and the leading approach to pharmacokinetic modelling, non-linear mixed effects modelling. The key limitation of the present study was the size of the data set. Though the data set was large

and detailed in the world of clinical pharmacology, it was limited for the purpose of comparing the methods described. Because of the risks associated with remifentanil infusion at larger than usual doses, it was not possible to obtain data for external validation of the models. As a result, our approach to performance estimation may have produced optimistic estimates of method accuracy. We sought, however, to maximize the amount of data available for training because prior work had demonstrated the importance of a large number of training instances in building ANN pharmacodynamic models.[3,24,25] An alternative approach to performance estimation would have been leave-one-out cross-validation.[26] In this circumstance, the standard modelling approach (non-linear mixed effects modelling) would serve as a useful point of reference.

Parameter selection may have influenced the relative performance of the methods in the present study. When multiple models were created using each method, the models appeared similar in structure and accuracy. The reduction of variance in individual predictions was a key benefit of the ensemble approach, which combined the predictions of multiple methods and models.

In clinical settings, remifentanil is usually infused at a rate of 0.25 μg/kg per min.[27] In contrast, the infusion rates of remifentanil in the present study ranged from 1 to 8 μg/kg per min in order clearly to assess the dose–response relationship. The dosing scheme in the present study was determined on the basis of volunteer safety. It was expected that young volunteers were more likely than middle-aged or elderly volunteers to maintain stable vital signs during a high-dose infusion of remifentanil. Remifentanil was, therefore, infused at rates of 1 – 8 μg/kg per min for 20 min in the young subjects to obtain as wide a distribution of concentrations and effects as possible. For the sake of safety, however, the rate and duration of infusion and, hence, the total dose of remifentanil were reduced in middle-aged and elderly volunteers, using the objective criteria of real-time SEF_{95}.[5]

In conclusion, new methods of pharmacokinetic modelling are emerging as the value of predicting individual patient blood concentrations becomes clear.[3] Previous studies have established the accuracy of ANNs in predicting blood concentrations of several pharmacotherapeutic agents. Previous work demonstrated the clinical value of predicting the blood concentrations of individual patients for the anaesthetic remifentanil using pharmacodynamic models.[3] The pharmacokinetic models that generate individual predictions require improvement and studies comparing multiple methods, such as SVMs and ensemble methods, using the same pharmacokinetic data are scarce. The present study compared the accuracy of an increasingly useful machine learning method, the SVM, to that of the ANN and the non-linear mixed effects model for pharmacokinetic modelling. It also examined the accuracy of simple averaging ensembles that combined the predictions of multiple methods. These basic ensembles appeared to temper the size of larger residuals, pulling outliers back into the distribution. This feature may be clinically useful when models are applied to predict individual blood concentrations of pharmacotherapeutic agents.

Acknowledgement

This study was supported by a 2006 Inje University Special Research Grant.

Conflicts of interest

The authors had no conflicts of interest to declare in relation to this article.

• Received for publication 27 March 2009 • Accepted subject to revision 2 June 2009
• Revised accepted 1 October 2009
Copyright © 2009 Field House Publishing LLP

References

1 Chow HH, Tolle KM, Roe DJ, *et al*: Application of neural networks to population pharmacokinetic data analysis. *J Pharm Sci* 1997; **86:** 840 – 845.
2 Tolle KM, Chen H, Chow HH: Estimating drug/plasma concentration levels by applying neural networks to pharmacokinetic data sets. *Decision Support Systems* 2000; **30:** 139 – 151.
3 Kang SH, Poynton MR, Kim KM, *et al*: Population pharmacokinetic and pharmacodynamic models of remifentanil in healthy volunteers using artificial neural network analysis. *Br J Clin Pharmacol* 2007; **64:** 3 – 13.
4 Durisová M, Dedík L: New mathematical methods in pharmacokinetic modeling. *Basic Clin Pharmacol Toxicol* 2005; **96:** 335 – 342.
5 Noh GJ, Kim KM, Jeong YB, *et al*: Electroencephalographic approximate entropy changes in healthy volunteers during remifentanil infusion. *Anesthesiology* 2006; **104:** 921 – 932.
6 Beal S, Sheiner LB: *NONMEM User's Guides*. San Francisco: NONMEM Project Group, University of California, 1992.
7 Jonsson EN, Karlsson MO: Xpose – an S-PLUS based population pharmacokinetic/pharmacodynamic model building aid for NONMEM. *Comput Methods Programs Biomed* 1999; **58:** 51 – 64.
8 Shapiro SW, Wilk M, Chen HJ: A comparative study of various tests for normality. *J Am Stat Assoc* 1968; **63:** 1343 – 1373.
9. Parke J, Holford NH, Charles BG: A procedure for generating bootstrap samples for the validation of nonlinear mixed-effects population models. *Comput Methods Programs Biomed* 1999; **59:** 19 – 29.
10 Kataria BK, Ved SA, Nicodemus HF, *et al*: The pharmacokinetics of propofol in children using three different data analysis approaches. *Anesthesiology* 1994; **80:** 104 – 122.
11 Matignon R: *Neural Network Modeling Using SAS Enterprise Miner*. Bloomington: AuthorHouse, 2005.
12 Smola AJ, Schölkopf B: A tutorial on support vector regression. *Stat Comput* 2004; **14:** 199 – 222.
13 Mandema JW: Population pharmacokinetics and pharmacodynamics. In: *Pharmacokinetics: Regulatory, Industrial, Academic Perspectives*, 2nd edn (Welling P, ed). New York: Informa Healthcare, 1995; pp 411 – 450.
14 Pinheiro JC, Bates DM: *Model Building for Nonlinear Mixed-effects Models*. Technical Report 91, Madison: Department of Biostatistics, University of Wisconsin–Madison, 1995.
15 Brier ME, Zurada JM, Aronoff GR: Neural network predicted peak and trough gentamicin concentrations. *Pharm Res* 1995; **12:** 406 – 412.
16 Vapnik VN: *The Nature of Statistical Learning Theory*. New York: Springer, 1995.
17 Dreiseitl S, Ohno-Machado L, Kittler H, *et al*: A comparison of machine learning methods for the diagnosis of pigmented skin lesions. *J Biomed Inform* 2001; **34:** 28 – 36.
18 Güler NF, Koçer S: Use of support vector machines and neural network in diagnosis of neuromuscular disorders. *J Med Syst* 2005; **29:** 271 – 284.
19 Liu HX, Zhang RS, Luan F, *et al*: Diagnosing breast cancer based on support vector machines. *J Chem Inf Comput Sci* 2003; **43:** 900 – 907.
20 Wei L, Yang Y, Nishikawa RM, *et al*: A study on several machine-learning methods for classification of malignant and benign clustered microcalcifications. *IEEE Trans Med Imaging* 2005; **24:** 371 – 380.
21 Hornik K, Stinchcombe M, White H: Multilayer feedforward networks are universal approximations. *Neural Netw* 1989; **2:** 359 – 366.
22 Sharkey AJ: On combining artificial neural nets. *Connection Sci* 1996; **8:** 299 – 314.
23 Perrone M: *Improving Regression Estimation: Averaging Methods for Variance Reduction with Extensions to General Convex Measure Optimization*. PhD thesis, Rhode Island: Brown University, 1993.
24 Chetchotsak D, Twomey JM: Combining neural networks for function approximation under conditions of sparse data: the biased regression approach. *Int J General Systems* 2007; **36**: 479 – 499.
25 Twomey JM, Smith AE: Performance measures, consistency, and power for artificial neural network models. *J Math Comput Model* 1995; **21:** 243 – 258.
26 Hastie T, Tibshirani R, Friedman J: *The Elements of Statistical Learning*. New York: Springer, 2001; pp 214 – 217.
27 Mastronardi P, Dellacasa P: Observational study on the use of remifentanil in general anesthesia. Drug utilisation research. *Minerva Anestesiol* 2004; **70:** 605 – 616.

Author's address for correspondence
Dr Sung Hong Kang
School of Health Administration, Inje University, 607 Obang-Dong, Kimhae,
Gyeongnam 621-749, Republic of Korea.
E-mail: hcmkang@gmail.com