December 2005

Pharmacokinetic Data Mining for Managing Clinical Use of Vancomycin

Paul Hu
University of Utah

Tsang-Hsiang Cheng
Southern Taiwan Univ. of Technology

Chih-Ping Wei
National Sun Yat-sen University

Pharmacokinetic Data Mining for Managing Clinical Use of Vancomycin

Paul Jen-Hwa Hu
Department of Accounting and Information Systems
University of Utah
actph@business.utah.edu

Tsang-Hsiang Cheng
Dept. of Business Admin.
Southern Taiwan Univ. of Technology
cts@mail.stut.edu.tw

Chih-Ping Wei
Department of Information Management
National Sun Yat-sen Univ.
cwei@mis.nsysu.edu.tw

Abstract

Drug-related problems have become a growing concern in clinical medicine. Particularly prevalent are problems surrounding sub- and over-therapeutic doses of high-alert medications. In this study, we use a model-tree based regression (namely M5) and an SVM technique to build systems that predict the regimen adequacy of vancomycin, a glycopeptide antimicrobial antibiotic effective for Gram-positive bacterial infections that has a narrow therapeutic index and considerable toxicity and adverse effects. We evaluate each system's predictions of the patient's peak and trough concentrations resulting from a vancomycin regimen, using a total of 1,099 clinical cases collected from a major tertiary medical center in southern Taiwan. We also examine the use of Bagging for enhancing the prediction accuracy of the respective systems and include in our evaluation a prevailing one-compartment model for performance benchmark purposes. Overall, our analysis results show that both M5 and SVM are significantly more accurate in predicting peak and/or trough concentrations than the one-compartment model, which approximates the current practice at the studied medical center. In addition, M5 appears to benefit considerably from Bagging, which also has a positive effect on SVM but seemingly to a lesser extent. Taken together, our findings suggest promising utilities of supervised learning techniques for improving the clinical use of vancomycin and similar high-alerting drugs, thus complementing clinicians' current practices.

Keywords: Decision Support in Medicine, Supervised Learning, Pharmacokinetic Data Mining, Management of Clinical Use of Vancomycin, Model-Tree Based Regression, Support Vector Machines, Bagging

1. Introduction

The fast-growing pharmaceutical drugs available for patient care have made the management of their clinical use increasingly important. Inadequate and erroneous medications and the resulting adverse patient reactions have been recognized as medical malpractices by the American Society Health System of Pharmacists since 1995. Problems surrounding inadequate medications and associated effects not only adversely affect the patient's health and well-being but also demand enormous resources for corrections. In the United States alone, the cost of clinical drug-related problems amounted to $76.6 billion in 1995 (Johnson and Bootman 1995) and may exceed $177.4 billion in the new millennium (Ernst and Grizzle 2001). Of particularly alarming prevalence are problems pertinent to sub- or over-therapeutic doses of high-alert medications, which have been identified by the Joint Commission on Accreditation of Healthcare Organizations as a major concern in the prevention of health care errors (Shirley et al. 2003).

Sub- and over-therapeutic dose problems result from inadequate regimen decision making[1] by clinicians and often have profound implications to clinical efficacy and safety. Such decision making is particularly critical when involving therapeutic drugs that have a narrow therapeutic index and high toxicity (impairments) for patients. A good case in point is vancomycin, a glycopeptide antimicrobial antibiotic effective for Gram-positive bacterial infections. Because of its advantages in bactericidal effects, sustained concentrations in serum achieved, and limited bacterial resistance (Michel and Gutmann 1997; Vincent et al. 1995), vancomycin has been widely used for treating methicillin-resistant staphylococcal infections. This particular antibiotic is a popular choice among clinicians for treating staphylococcal aureus and methicillin-resistant staphylococcal epidermidis (Michel and Gutmann 1997; Vincent et al. 1995). Vancomycin is also commonly used for curing patients suffering from streptococcus phylogenies, streptococcus pneumoniae, enterococci, corynebateririum, and clostridium difficile as well as for patients allergic to penicillin or cephalosporin.

Clinically, measures for reducing the likelihood of a clinician's prescribing inadequate regimens are more effective and appealing for both patient care and economic considerations than retrospective methods for regimen corrections or adjustments. Most current practices greatly depend on the use of the nomogram or pharmacokinetic analysis for estimating or evaluating vancomycin regimens; however, these methods may not be effective. In general, the nomogram provides clinicians with a simplistic baseline for (initial) regimen estimations but does not allow predictions of a regimen's adequacy with respect to the target (normal) range of peak and/or trough concentration[2]. Meanwhile, pharmacokinetic analysis uses statistical models derived from clinical pharmacokinetic data of patients. As Chow et al. (1997) summarized, building a population pharmacokinetic model requires understanding and selection of various mathematical/statistical models that include a pharmacokinetic structure model for relating dosage, sampling time, and pharmacokinetic parameters to plasma drug levels; regression models describing the relationships between patient characteristics and the pharmacokinetic parameters; a population model for inter-subject variability; and a variance model for random residual variation in the data (intra-patient variability). Moreover, the effectiveness of a pharmacokinetic model is greatly affected by the characteristics of the population pharmacokinetic data. Tolle et al. (2000) suggested that repeated measures (i.e., a patient contributing a series of clinical cases), imbalance (i.e., unequal number of clinical cases per patient), and confounding (a patient receiving specific dosages based on that patient's prior reactions) problems inherent to the population pharmacokinetic data make it difficult to develop statistical analysis tools for predicting pharmacokinetic parameters.

Thus, the motivation of this study is to investigate promising AI-based supervised learning techniques for pharmacokinetic analysis. Such AI-based pharmacokinetic analysis systems predict the patient's peak/trough concentration levels and allow clinicians to assess regimen adequacy of vancomycin. From the perspective of clinical practice, they can enhance the efficacy of primitive regimen estimations based on the nomogram and, at the same time, complement the pharmacokinetic analysis. For example, a clinician could estimate an initial

[1] Typically, a clinician's regimen decision includes the daily dosage and the time interval between administrations for the therapeutic drug in question.

[2] Peak value refers to the concentration in serum detected "immediately" after the patient receives the medication, whereas trough value denotes the concentration detected "before" the next administration. In the case of vancomycin, peak value is often examined 30–60 minutes after the administration, and the trough value is taken right before the next administration. Although not universally agreed on, a peak concentration between 20μg/ml and 40μg/ml and a trough concentration below 10μg/ml are commonly considered to be "appropriate" for vancomycin by clinicians (Leader et al. 1995).

regimen using the nomogram and then use the prediction system to verify its clinical adequacy.

The specific supervised learning techniques examined in this study are the model-tree based regression (M5) and SVM. Our choice of techniques was mainly based on the interpretability of their prediction results and computational efficiency. We empirically evaluated the accuracy of each system using mean absolute error, calculated as the absolute value of the difference between the predicted and known concentration levels. For performance benchmark purposes, our evaluation included a prevailing one-compartment pharmacokinetic model which approximates the current practice in the studied medical center. We comparatively evaluated the systems independently as well as in conjunction with the use of Bagging. This study used a total of 1,099 vancomycin regimens clinically administered January 2000 and July 2004 at a major tertiary medical center in southern Taiwan. Overall, our analysis shows that both M5 and SVM are significantly more accurate in predicting the patent's peak and trough concentrations (resulting from a vancomycin regime) than the benchmark one-compartment model. Judged by the predicted peak and trough concentrations, M5 is more effective than SVM and appears to benefit considerably from Bagging.

The remainder of the paper is organized as follows. Section 2 provides an overview of the research background (vancomycin and its clinical use) and reviews relevant previous research to highlight our motivation. Section 3 describes our data collection and the particular supervised learning techniques examined. Section 4 details the design of the respective prediction systems and discuses our evaluation design, followed by important evaluation results in Section 5. We conclude in Section 6 with a summary and discussions of some future research directions.

2. Background Overview and Literature Review

In this section, we provide an overview of vancomycin and its clinical use, and then review relevant previous research.

2.1 Vancomycin and Its Clinical Use

Proper use of vancomycin in patient care has been a challenge to clinicians. This antibiotic is a popular choice among clinicians for treating Gram-positive bacterial infections, staphylococcal aureus, and methicillin-resistant staphylococcal infections primarily because of its advantageous bactericidal effects, sustained serum concentrations, and limited bacterial resistance. However, vancomycin has a narrow therapeutic index and, when inappropriately prescribed, can result in severe lasting adverse effects on the patient. Specifically, insufficient (i.e., sub-therapeutic) doses yield limited therapeutic effects and allow bacteria to become increasingly drug resistant. On the other hand, excessive (i.e., over-therapeutic) doses produce harmful deposits of vancomycin, which can lead to severe renal function impairments, kidney poisoning, ear poisoning, or combinations of such problems. The described adverse effects become even greater clinical threats when patients have impaired renal functions[3] or are receiving other nephrotoxic drugs (Pea et al. 2002). According to the pharmacovigilance issued by the National Reporting System of Adverse Drug Reactions (http://adr.doh.gov.tw), vancomycin consistently has been identified as one of the top three adverse effect-producing pharmaceutical drugs in Taiwan between 1998 and 2003.

[3] In general, patients can discharge 80%–90% of the injected vancomycin (in its original form) through urination. Patients who suffer from deteriorated renal function have a lower discharge rate and a longer half-life decay cycle and therefore are more likely to develop excessive vancomycin deposits, which further damages their renal function and can lead to kidney poisoning.

To cope with the constricted therapeutic index and potential toxicity (and impairments) of vancomycin, many clinicians follow general guidelines for prescribing this antibiotic. A particularly popular method is the nomogram (Moellering et al. 1981) for estimating adequate vancomycin regimens based on creatinine clearance (CLCR), which is a function of the patient's age, gender, weight, and serum creatinine (SCR, a common renal function index). In general, clinicians should target an average vancomycin concentration at the 15 μg/ml level (Moellering et al. 1981); accordingly, the nomogram was developed based on the clinical efficacy observed from a large patient population. A similar clinical method uses an initial vancomycin regimen of 8 mg/kg and adjusts the time interval between administrations based on the estimated CLCR of the patient (Lake and Peterson 1988). Clinical guidelines drawn from the nomogram are intuitive and easy to use and can provide a reasonable baseline for (initial) vancomycin regimen estimations (Matzke et al. 1985). However, the nomogram is primitive and therefore may not be sophisticated or effective enough to meet the clinical need of individual patients. As a result, clinicians usually have to supplement the nomogram with their own clinical experiences and patient condition assessments to adjust recommendations by the nomogram.

2.2 Literature Review and Motivation

Prior research has suggested the use of pharmacokinetic analysis to better support clinical use of vancomycin. A pharmacokinetic model represents a mathematical scheme that depicts complex physiologic spaces or processes. Accurate pharmacokinetic modeling is crucial for determining the elimination (or discharge) rate of an administered drug. Both one- and two-compartment pharmacokinetic models have been developed. An administered drug initially is distributed into a central compartment before diffusing into the peripheral compartment. When dealing with drugs that equilibrate with the tissue compartment quickly, clinicians, for practical purposes, usually choose a one-compartment model that involves only one volume term; i.e., volume of distribution. In a one-compartment model, the peak and trough concentration levels can be estimated using the following equations (Winter 2003):

$$CP_{SS(max)} \text{ (predicted peak concentration)} = \frac{\frac{S \times F \times Dosage}{V_d}}{1 - e^{-K_d \times \tau}} \text{ and}$$

$$CP_{SS(min)} \text{ (predicted trough concentration)} = CP_{SS(max)} \times (e^{-K_d \times \tau})$$

where S (salt form parameter) is assumed to be 1 for vancomycin, F (bioavailability parameter) is assumed to be 1 for vancomycin because orally administered vancomycin is poorly absorbed (such as less than 5%) and parenteral administrations are necessitated, *Dosage* is the dosage (in mg), V_d (volume of distribution) = 0.17×age + 0.22×weight + 15, K_d (elimination rate constant) depicts the rate at which vancomycin is discharged from the patient body and is estimated as $\frac{\text{clearance for vancomycin}}{V_d}$ where clearance for vancomycin is estimated as creatinine clearance (CL_{CR}), τ is the dosing interval (in hour), and $CL_{CR} =$

$$\begin{cases} \frac{(140 - \text{age}) \times \text{weight}}{72 \times \text{SCR}} & \text{if patient is male} \\ \frac{(140 - \text{age}) \times \text{weight}}{72 \times \text{SCR}} \times 0.85 & \text{if patient is female.} \end{cases}$$

When a drug equilibrates slowly with the peripheral tissues, clinicians often describe its distribution in the human body by a two-compartment pharmacokinetic model. Overall, pharmacokinetic analyses based on two-compartment models can considerably reduce biases and therefore are more likely to improve the prediction accuracy for peak and trough concentrations (Jelliffe et al. 1993; Jelliffe et al. 1994). To derive reliable estimates for pharmacokinetic parameters commonly required by a two-compartment model, additional concentration-in-serum values are needed to fully characterize the patient's serum concentration–time profile. When using a nonlinear least squares regression analysis of a two-compartment model, a minimum of four concentrations for each clinical case included in the population pharmacokinetic dataset is required to construct the model. These stringent requirements greatly reduce the clinical practicality and value of two-compartment models, which therefore are not commonly used by clinicians in their patient care (Hurst et al. 1990). Previous research has also examined the Bayesian approach to combining both population and patient-specific information (such as serum-level data) to predict serum concentrations. Clinical evaluations show Bayesian methods capable of achieving greater predictive power than conventional one- or two-compartment models (Hurst et al. 1990; Jelliffe et al. 1994). However, as with two-compartment models, Bayesian methods require considerably more serum samples from patients, which seriously limit their clinical practicality and value.

Whereas pharmacokinetic analyses based on one-compartment models are more feasible or affordable clinically; however, their accuracy in predicting vancomycin peak and trough concentrations is constrained (Jelliffe et al. 1993; Jelliffe et al. 1994). To improve accuracy of concentration prediction, AI-based supervised learning techniques are promising and have been examined by prior research. A review of extant literature suggests the prevalence of neural network. In one investigation of population pharmacokinetic data, the neural network appeared capable of capturing the relationships between plasma drug levels and patient-related prognostic factors from routinely collected pharmacokinetic data (Chow et al. 1997). Tolle et al. (2000) compared the effectiveness of a feed-forward backpropagation neural network and NONMEM (popular software based on a one-compartment pharmacokinetic model) for predicting the serum concentration levels of administered tobramycin in patients suffering from pediatric cystic fibrosis and hemotologic-oncologic disorders. According to the evaluation results, the neural network approach appears to be a cost-effective approach that provides clinicians and pharmacists with timely analysis results that complement those of NONMEM. In another related study (Yamamura et al. 2003), a neural network was developed to identify intensive-care patients whose drug concentrations were likely to fall below therapeutic levels. The evaluation results show that the neural network outperformed multivariate logistic regression analysis.

While demonstrating encouraging clinical efficacy, neural networks are computationally demanding and often offer limited (if any) interpretability of the prediction results. Alternative supervised learning techniques are largely ignored by previous research. For a perspective of both research and clinical practice, it is important to compare the effectiveness of other promising supervised learning techniques by examining their predictions of the patient's peak and/or trough concentrations. Such evaluations are particularly important for vancomycin and other antibiotics with a narrow therapeutic index and significant toxicity or probable impairments. In the current study, we used a model-tree based regression (specifically, M5) (Quinlan 1992) and a SVM technique (Smola and Schölkopf 2004; Vapnik 1995) to develop prediction systems for estimating vancomycin peak/trough concentration levels and evaluated their respective clinical efficacy using data collected from a tertiary medical center in southern Taiwan.

3. Data and Investigated Supervised Learning Techniques

In this section, we describe our data collection and the particular AI-based supervised learning techniques investigated; i.e., M5 and SVM as well as Bagging.

3.1 Data Collection

We collected a total of 1,099 clinical vancomycin cases administered between January 2000 and December 2004 at a tertiary medical center in southern Taiwan. Typically, clinicians at this medical center use the nomogram to estimate initial vancomycin regimens and depend on a one-compartment pharmacokinetic model for making subsequent regimen adjustments. Each case consisted of the patent's gender, age, weight, SCR, the vancomycin regimen he or she received (i.e., dosage and time interval between administrations), and the resulting peak and trough concentrations detected by TDM. Clinically, the attending clinician used the recorded peak and trough concentrations to evaluate the adequacy of the regimen previously prescribed and make regimen adjustments accordingly. Table 1 summarizes the specific variables included in each vancomycin case and their descriptive statistics.

Table 1: Summary of Variables of Clinical Cases and Descriptive Statistics

Variables	Range	Descriptive Statistics
Gender	Male or Female	Male: 739 (67.24%); Female: 360 (32.76%)
Age (in years)	18 to 96	μ (mean) = 61.08; σ (standard deviation)= 16.26
Weight (in kilograms)	31.1 to 121	μ = 63.19; σ = 13.90
Serum creatinine (SCR)	0.22 to 11.3	μ = 1.55; σ = 1.54
Dosage (in mg)	50 to 1500	μ = 706.82; σ = 253.60
Dosing interval (in hours)	6, 8, 12, 24, 48, 72, 96, or 168	6 hours: 205 (18.65%) 8 hours: 200 (18.20%) 12 hours: 397 (36.12%) 24 hours: 200 (18.20%) 48 hours: 73 (6.64%) 72 hours: 20(1.82%) 96 hours: 2 (0.18%) 168 hours: 2 (0.18%)
Trough concentration level	0.9 to 99.5	μ = 18.03; σ = 12.56
Peak concentration level	8.8 to 220.4	μ = 34.54; σ = 16.79

The specific inputs to each investigated prediction system were gender, age, weight, serum creatinine (SCR), dosage, and dosing interval. In this study, we did not include the variables of the time of blood drawn for examining peak and trough concentrations because, for each clinical case collected, the patient's peak concentration was obtained approximately 30 minutes after the administration and his or her trough concentration was recorded immediately before the next administration (i.e., the time of blood drawn for examining the trough concentration was equal to the dosing interval of the target case). We used a 1-of-N coding scheme to represent dosing interval because its values are of ordinal rather than continuous nature. Together, each prediction problem comprised of 13 independent input variables. In our evaluation, each system predicted peak and trough concentrations independently; i.e., two dependent output variables.

3.2 Investigated Supervised Learning Techniques: M5, SVM, and Bagging

In a nutshell, the use of a supervised learning technique for prediction is to build an automated prediction model that captures important relationships between a set of input variables and a dependent variable, commonly of numeric value. That is, a selected learning technique uses a set of training instances (with known values for independent and dependent variables) to build a prediction model. The resulting model is then applied for predicting the dependent variable of a new (unseen) instance, based on its input variable values. Different techniques have been proposed and empirically examined in various application contexts, including standard regression, neural network, model-tree-based regression, and SVM. In general, standard regression may not be robustly potent because its prediction function representation explicitly demands linear relationships among the independent variables. Neural network has been shown to be reasonably effective and robust across different applications but it is computationally demanding and provides limited, if any, interpretability. Model-tree based regression and SVM techniques are also popular and have demonstrated satisfactory effectiveness in different prediction applications. Accordingly, we adopted a model-tree based regression technique (namely, M5) and a SVM technique to build prediction models for the patient's peak and trough concentrations resulting from a particular vancomycin regimen.

Quinlan (1992) proposed M5, a model-tree based regression technique for prediction analysis. M5 combines a conventional decision tree with the possibility of linear regression functions at leaf nodes. Structurally, a model tree resembles a decision tree but has linear regression functions at its leaf nodes rather than discrete (output) classes common to decision trees. Constructing a model tree involves tree building and tree pruning (Quinlan 1992; Wang and Witten 1997). In the tree building stage, a decision tree induction algorithm is applied to build the tree. Unlike such classical decision tree induction algorithms as ID3 (Quinlan 1986) and C4.5 (Quinlan 1993) that maximize the information gain (or ratio) at each intermediate node, M5 employs an attribute-selection criterion that minimizes the intra-subset variation of the dependent variable at each branch. This selection criterion measures the error of a target node using the dependent variable's standard deviation of those instances pertaining to the node and calculates the expected standard deviation reduction that results from splitting the target node according to a chosen attribute. As a result, M5 selects the attribute that yields the greatest standard deviation reduction and uses it for branching the target node in the decision tree. For each child node, the described attribute selection and tree building process continues recursively until all the instances that reach the node show trivial variations or contains only a small number of instances.

In the tree pruning stage, M5 uses an estimate for the expected error at each node *T*. First, the absolute difference (error) between the predicted and actual value of the dependent variable is calculated for each training instance and then averaged across all training instances that reach *T*. The resulting average error is likely to underestimate the expected error for unseen instances. To compensate, the average error is multiplied by $\frac{n+v}{n-v}$, where n is the number of training examples pertaining to *T* and v is the number of independent variables used for representing the prediction model at *T*. M5 computes a standard linear regression model for each intermediate node of the un-pruned tree. Each resulting linear model can be simplified by removing selected independent variables, thus minimizing the estimated error calculated using the above multiplication factor. Removal of an independent variable reduces the multiplication factor, which may offset the inevitable increase in the average error over all training instances. Such independent variable removals proceed following a greedy strategy; i.e., one by one until the compensated error estimate decreases. Once a linear regression

model is determined for each intermediate node, the tree is pruned reversed starting from the leaf nodes, terminated when the expected estimate of error decreases.

Developed by Vapnik (1995), Support Vector Machine (SVM) creates from a set of training instances a classification function for the targeted classification problem or a general regression function for the intended prediction analysis. SVM is applied to the regression setting maintaining all the features of the maximal margin concept. The capacity of the system is controlled by the regularization parameter *C*. The regression setting is about finding a function that approximates the mapping between the input domain and the real-number domain. The difference between the actual and the estimated output of a training instance is the residual for the instance, which, in turn, shows the accuracy of the mapping. Fundamentally, SVM minimizes large residuals but tolerates small residuals, and determines their balancing based on *ε-insensitive loss function*.

In *ε*-SVM regression, the objective is to find a function $f(x)$ that has at most ε deviation from the actually output y_i for all the training instances. SVM strives for minimizing the loss function as well as the Euclidean norm of linear parameters ($||w||^2$), which resembles a problem of convex optimization. A soft-margin concept is used by introducing non-negative slack variables, ξ and ξ^2, that measure the deviation of training instances outside the *ε*-insensitive tube. Thus, SVM can be formulated as the minimization of the following function.

$$\frac{1}{2}(||w||^2) + C \sum_{i=1}^{n}(\xi_i+\xi_i^*)$$

subject to

$$y_i - w^t x_i - b \le \varepsilon + \xi_i$$
$$-y_i + w^t x_i + b \le \varepsilon + \xi_i^*$$
$$\xi_i^*, \xi_i \ge 0$$

This optimization problem can be solved as a quadratic programming problem. The constant $C > 0$ determines the tradeoff between empirical error and the structural error (i.e., error on the training instances versus generalization error). The $\frac{1}{2}(||w||^2)$ term controls the complexity of the regression function.

We also investigated the use of the ensemble approach for enhancing the predictive power of M5 and SVM. This approach creates a learning setting in which a finite set of prediction models (hereafter, base predictors) are constructed (i.e., trained) and used together to arrive at an overall prediction for a new (unseen) instance. Prevailing ensemble techniques include Bagging (Bauer and Kohavi 1999; Breiman 1996; Quinlan 1996; Zhou et al. 2002) and Boosting (Bauer and Kohavi 1999; Freund 1996; Freund and Schapire 1996; Quinlan 1996; Schapire 1990). Results from several empirical studies showed that Bagging is capable of reducing the classification variances significantly, and that the effectiveness of Boosting may be unstable (Quinlan 1996; Zhou et al. 2002). Therefore, we chose Bagging to improve the performance of the respective prediction models primarily because of its relative stability and the favorable empirical support for its prediction accuracy enhancements.

Bagging employs bootstrap sampling to generate multiple training data sets from the original overall training data set. Each resulting training data set is the used to construct a base

predictor. In each iteration, Bagging randomly samples the original overall training data set with replacements to create a distinct data set that has the same number of training instances as in the original data set. Each resulting data set provides training cases to constructing a prediction model based on a specific learning technique; in our case, M5 or SVM. This process terminates after a pre-specified number of iterations. To predict the output value for a new (unseen) instance, an overall prediction is made using the averaged predictions made by all base predictors constructed.

4. Prediction System Development and Evaluation Design

In this section, we describe the design of the prediction systems under evaluation, detail our evaluation design (including the procedure and performance measurements), and discuss our parameter tuning experiments required by the Bagging-based extensions.

4.1 Design of Prediction Systems

We used Weka open-source machine learning software (www.cs.waikato.ac.nz/ml/weka/) to construct prediction systems based on M5[4] and SVM. We extended each investigated system using Bagging, hereafter referred to as Bagged M5 and Bagged SVM. The use of both Bagged M5 and Bagged SVM requires an additional parameter (i.e., number of iterations) that corresponds to the number of base predictors to be constructed. We conducted parameter-tuning experiments to experimentally determine an optimal number of iterations for each bagged extension (of which details are discussed in Section 4.3).

4.2 Evaluation Design

We followed an 80/20 training/testing strategy, hereby randomly dividing the overall cases into two subsets: training (80% of the overall cases) and testing (20%). We used the training cases to construct the respective prediction systems, each of which was then evaluated using the testing cases. To reduce potential biases from a single randomization, we performed the randomization 30 times, each involving different training and testing subsets. In this study, we report system performance using the average effectiveness across the 30 trials examined. Specifically, we evaluated each prediction system in terms of prediction accuracy that was measured using the average of absolute differences between predicted and actual peak (or trough) concentration levels; i.e., mean absolute error (MAE).

4.3 Parameter Tuning Experiments for Bagged M5 and Bagged SVM

For each bagged extension, we experimentally determined an optimal number of iterations by performing a series of parameter-tuning experiments. Following the evaluation design described in the previous section, we assessed prediction accuracy of each Bagging-based system over the range between 2 and 100 iterations. As shown in Figure 1, the MAE of peak concentration predictions by Bagged M5 decreased considerably over the range of 2 to 30 iterations and then leveled off. The bagging-induced accuracy improvement to SVM also appeared relatively noticeable between 2 and 30, but to a much less extent than that of M5. Analysis of the standard deviation of absolute errors showed similar patterns; both systems exhibiting considerable reductions in MAE between 2 and 30 iterations and becoming stable beyond 30 iterations.

[4] In WeKa, this technique is named M5P rather than M5.

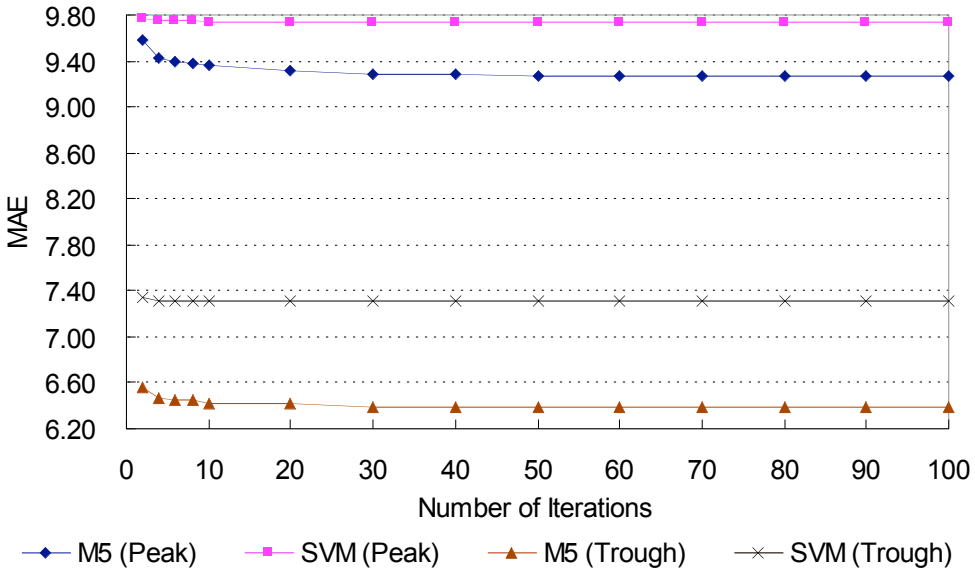

Figure 1: MAE of Concentration Predictions

Similar patterns were observed with the trough concentration predictions. As shown in Figure 1, the MAE of trough concentration predictions by Bagged M5 decreased considerably between 2 to 30 iterations and leveled off beyond this range. The prediction accuracy of SVM appeared to increase with the use of Bagging in the range of 2 and 30 iterations and then become stable. The bagging-enabled accuracy improvement was considerably larger with M5 than with SVM. Analysis of the standard deviation of absolute errors suggested that systems experienced more prominent reductions in MAE between 2 and 30 iterations than beyond that range. Overall, our tuning results suggest the use of 30 iterations to be appropriate. We therefore set the number of iterations at 30 for both bagged systems, each of which had 30 base predictors.

Bagged SVM appeared to be relatively insensitive to the number of iterations. The differential prediction improvements resulting from the use of bagging may in part be explained by the learning strategy used by the respective systems. In general, M5 adopts a divide-and-conquer strategy in the model tree building process and, thus, marginal variation in training datasets may generate noticeably different model trees characterized by their (tree) structures and the specific regression functions in leaf nodes. As mentioned, Bagging uses a random sampling process to generate multiple data subsets and, in the case of M5, is likely to result in different base predictors from the iterations examined. The non-identical base predictors of bagged M5 collectively could result in noticeable predictive accuracy improvements. The relative insensitivity of SVM to Bagging might also be reasoned by its learning strategy; that is, SVM does not employ a similar divide-and-conquer strategy and therefore its incorporation with bagging would likely generate a set of highly homogeneous base predictors. Consequently, Bagged SVM is not likely to benefit considerably from this particular ensemble approach.

5. Comparative Evaluation Results and Discussions

In this section, we report results from our comparative evaluations of M5, SVM, and their bagging extensions (based on iterations of 30). The one-compartment pharmacokinetic model depicted in Section 2.2 was included in our evaluation for performance benchmark purposes. Overall, as shown in Table 2, our comparative analysis showed both M5 and SVM significantly outperforming the benchmark one-compartment model in predicting the

patient's peak and trough concentrations. Particularly, the use of M5 or SVM reduced MAE of peak concentration by nearly 50% and resulted in an even greater reduction in the MAE of trough concentration. The dispersion of prediction errors for peak and trough concentrations (measured by the standard deviation of absolute errors) also decreased when using M5 or SVM as opposed to the one-compartment model. Judged by the observed mean absolute errors, M5 appeared to benefit considerably from Bagging of which effect on SVM seems marginal at best.

Table 2: Overall Evaluation Results

	Peak Concentration		Trough Concentration	
	MAE	σ(AE)*	MAE	σ(AE)
One-Compartment Model	17.277	16.388	18.021	12.484
M5	9.639	12.341	6.489	6.772
SVM	9.731	13.164	7.321	8.437
Bagged M5†	9.281	12.126	6.393	6.640
Bagged SVM†	9.740	13.154	7.314	8.426

†: Number of iterations = 30

*: Standard deviation of absolute errors

We performed a series of t-tests to assess the statistical significance of the differences in peak-concentration prediction accuracy among the investigated systems, including the benchmark one-compartment model. As summarized in Table 3, the prediction accuracy of the one-compartment model was significantly lower than that of M5, SVM, or their respective bagging extensions, as suggested by a *p*-value less than 0.01. M5 appeared more effective than SVM but the difference was not significant statistically. The use of Bagging significantly improved the prediction accuracy of M5 but not that of SVM. In effect, Bagged M5 appeared the most effective for predicting the patient's peak concentration, significantly outperforming all the other systems investigated.

Table 3: Comparative Analysis of Peak-Concentration Predictions by Investigated Systems

	One-Compartment Model	M5	SVM	Bagged M5	Bagged SVM
One-Compartment Model					
M5	0.000***				
SVM	0.000***	0.193			
Bagged M5	0.000***	0.000***	0.000***		
Bagged SVM	0.000***	0.147	0.122	0.000***	

***: Significant at $p < 0.01$ on a two-tailed t-test.

We performed similar t-tests on the trough-concentration predictions by the respective systems. As shown in Table 4, the one-compartment model was significantly less accurate than M5, SVM, and their respective bagging extensions; i.e., *p*-value less than 0.01. M5 was more effective than SVM and the difference was statistically significant, as suggested by a *p*-value less than 0.01. The use of Bagging significantly improved the prediction accuracy of M5 but not that of SVM. Bagged M5 was significantly more accurate in predicting the patient's trough concentration than Bagged SVM. Jointly, according to comparative analysis results, M5 is more effective than SVM and the one-compartment model in predicting the patient's peak and trough concentrations and is likely to be benefit considerably from Bagging. The use of M5 and its bagging extension is particularly appealing in light of the clinical importance of the patient's trough concentration that is more critical to clinicians' regimen decision-making than the peak concentration (Herman 2003).

Table 4: Comparative Analysis of Trough-Concentration Predictions by Investigated Systems

	One-Compartment Model	M5	SVM	Bagged M5	Bagged SVM
One-Compartment Model					
M5	0.000***				
SVM	0.000***	0.000***			
Bagged M5	0.000***	0.000***	0.000***		
Bagged SVM	0.000***	0.000***	0.136	0.000***	

***: Significant at $p < 0.01$ on a two-tailed t-test.

Overall, our evaluation results suggest the followings. First, prediction systems constructed using supervised learning techniques would offer considerable clinical value for assessing the adequacy of a vancomycin regimen. Based on our findings, both M5 and SVM are significantly more accurate in predicting patients' peak and trough concentrations than the one-compartment model that approximates the current practice. In turn, this suggests promising utilities of such supervised learning techniques for addressing the problems surrounding the clinical use of vancomycin. Second, our analysis shows that Bagging may enhance the accuracy of prediction systems based on supervised learning techniques, particularly those adopting a learning strategy inherently benefiting from the heterogeneous base predictors generated by Bagging. As highlighted in our study, the Bagging-enabled accuracy improvement is more prominent with M5 than with SVM because of its use of a divide-and-conquer strategy likely to benefit from the differences in base predictors. Both M5 and SVM are not demanding computationally and have shown reasonable robustness to the size of training dataset. These characteristics make their clinical use for predicting or assessing the adequacy of high-alerting medications increasingly appealing, as compared to pharmacokinetic models. Taken together, our analysis results suggest building prediction systems using M5 (or similar supervised learning techniques) as well as considering the use of Bagging to enhance its prediction accuracy.

6. Conclusion

Mitigating the drug-related problems in patient care is critical. Using promising supervised learning techniques, the current research responds to the challenge of managing the clinical use of vancomycin, a popular antibiotic effective for Gram-positive bacterial infections but having a narrow therapeutic index and considerable toxicity and adverse effects. Specifically, we used M5 and SVM to build systems for predicting the patient's peak and trough concentrations resulting from a vancomycin regimen. We extended each system using Bagging and included a prevailing one-compartment model for performance benchmark purposes. Our overall results are encouraging and suggest the efficacy and value of M5 and SVM techniques for managing the clinical use of vancomycin, with or without Bagging. According to our analysis, both systems achieve a prediction accuracy significantly higher than that by the one-compartment model that approximates the current clinical practice at the studied medical center. The reported systems can complement existing practices. For instance, a clinician can use our systems to verify the adequacy of an initial regimen suggested by the nomogram as well as to re-evaluate the adequacy of a regimen adjustment suggested by the pharmacokinetic analysis result, before administering the suggested regimen on the patient or performing a tedious therapeutic drug monitoring.

The current research can be extended in several directions. First, we considered all regimens homogeneous in this study; further research should distinguish between, for example, initial versus follow-up regimens. With this differentiation, we can further scrutinize and compare

the effectiveness of the reported systems in different regimen categories. Similarly, assessing the systems' performance for different patients is also important, particularly for those patients with normal versus deteriorated renal functions (or kidney diseases), because of vancomycin's considerable potential impairments to renal functions. Second, whereas we evaluated the systems using archival vancomycin cases (which captured the clinical practices at the studied medical center), our continued investigations target direct performance comparisons between the systems and the nomogram and the pharmacokinetic analysis, particularly that based on a two-compartment model. Conducting similar evaluations involving clinicians and vancomycin cases from other healthcare institutions is also desirable and can enhance the systems' clinical efficacy or increase the validity of our evaluation results. Third, evaluations of the reported systems can be expanded. Of particular importance is predicting the likelihood that a regimen will produce adverse effects, which are of great concern in clinicians' use of vancomycin. In addition, our reported approach, including the automated construction of prediction systems and evaluation designs, can be and should be applied to other high-alert medications.

Acknowledgment

This work was supported in part by College of Management, National Sun Yat-sen University under the grant CMNSYSU-CTP-2004-01.

References

Bauer, E. and Kohavi, R., "An Empirical Comparison of Voting Classification Algorithms: Bagging, Boosting, and Variants," *Machine Learning* (36:1-2), 1999, pp.105-142.

Breiman, L., "Bagging Predictors," *Machine Learning* (24:2), 1996, pp.123-140.

Chow, H., Tolle, K. M., Roe, D. J., Elsberry, C. and Chen, H., "Application of Neural Networks to Population Pharmacokinetic Data Analysis," *Journal of Pharmaceutical Sciences* (86:7), 1997, pp.840-845.

Ernst, F. R. and Grizzle, A. J., "Drug-Related Morbidity and Mortality: Updating the Cost-of-Illness Model," *Journal of the American Pharmacists Association* (41:2), 2001, pp.192-199.

Freund Y. and Schapire, R. E., "Experiments with a New Boosting Algorithm," *Proceedings of the Thirteenth National Conference on Machine Learning*, Morgan Kaumann, Bari, Italy, 1996, pp.148-156.

Freund, Y., "Boosting a Weak Learning Algorithm by Majority," *Information and Computation* (121:2), 1996, pp.256-285.

Herman, R., "Current Issues Related to Therapeutic Drug Monitoring: Vancomycin," *World of Drug Information* (14:1), 2003, pp.1-4, 6.

Hurst, A .K., Yoshinaga, M. A., Mitani, G. H., Foo, K. A., Jelliffe, R. W. and Harrison, E. C., "Application of a Bayesian Method to Monitor and Adjust Vancomycin Dosage Regimens," *Antimicrob Agents Chemother* (34:6), 1990, pp.1165-1171.

Jelliffe, R. W., Maire, P., Sattler, F., Gomis, P., and Tahani, B., "Adaptive Control of Drug Dosage Regimens: Basic Foundations, Relevant Issues, and Clinical Examples," *International Journal of Bio-medical Computing* (36:1-2), 1994, pp.1-23.

Jelliffe, R. W., Schumitzky, A., Van Guilder, M., Liu, M., Hu, L., Maire, P., Gomis, P., Barbaut, X., and Tahani, B., "Individualizing Drug Dosage Regimens: Roles of Population Pharmacokinetic and Dynamic Models, Bayesian Fitting, and Adaptive Control," *Therapeutic Drug Monitoring* (15:5), 1993, pp.380-393

Johnson, J. A. and Bootman, J. L., "Drug-Related Morbidity and Mortality: A Cost-of-Illness Model," *Archives of Internal Medicine* (155), 1995, pp.1949-1956.

Lake, K. D. and Peterson, C. D., "Evaluation of a Method for Initiating Vancomycin Therapy: Experience in 205 Patients," *Pharmacotherapy* (8), 1988, pp.284-286.

Leader, W. G., Chandler, M. H., and Castiglia, M., "Pharmacokinetic Optimisation of Vancomycin Therapy," *Clinical Pharmacokinetics* (28), 1995, pp.327-342.

Matzke, G. R., Kovarik, J. M., Rybak, M. J., and Boike, S. C., "Evaluation of the Vancomycin-Clearance: Creatinine-Clearance Relationship for Predicting Vancomycin Dosage," *Clinical Pharmacy* (4), 1985, pp.311-315.

Michel, M. and Gutmann, L., "Methicillin-Resistance Staphylococcus Aureus and Vancomycin-resisitant Enterococci: Therapeutic Realities and Possibilities," *The Lancet* (349), 1997, pp.1901-1906.

Moellering, R. C., Krogstad, D. J., and Greenblatt, D. J., "Vancomycin Therapy in Patients with Impaired Renal Function: A Nomogram for Dosage," *Annuals of Internal Medicine* (94), 1981, pp.343-346.

Pea, F., Bertolissi, M., Silvestre, A. D., Poz, D., Giordano, F., and Furlanut, M., "TDM Coupled with Bayesian Forecasting Should Be Considered an Invaluable Tool for Optimizing Vancomycin Daily Exposure in Unstable Critically Ill Patients," *International Journal of Antimicrobial Agents* (20), 2002, pp.326-332.

Quinlan, J. R., "Bagging, Boosting, and C4.5," *Proceedings of the Thirteenth National Conference on Artificial Intelligence (AAAI-96)*, Portland, OR, AAAI Press, Menlo Park, CA, 1996, pp.725-730.

Quinlan, J. R., "Induction of Decision Trees," *Machine Learning* (1:1), 1986, pp.81-106.

Quinlan, J. R., "Learning with Continuous Classes," *Proceedings of 5th Australian Joint Conference on Artificial Intelligence*, World Scientific, Singapore, 1992, pp.343-348.

Quinlan, J. R., *C4.5: Programs for Machine Learning*,Morgan Kaufmann, San Mateo, CA, 1993.

Schapire, R. E., "The Strength of Weak Learnability," *Machine Learning* (5:2), 1990, pp.197-227.

Shirley, J., Paine, J. D., Sarah, G., and Benator, J. D., "JCAHO Initiative Seeks to Improve Patient Safety," *Drug Benefit Trends* (15:1), 2003, pp.23-24.

Smola, A. J. and Schölkopf, B., "A Tutorial on Support Vector Regression," *Statistics and Computing* (14), 2004, pp.199-222.

Tolle, K. M., Chen, H., and Chow, H., "Estimating Drug/Plasma Concentration Levels by Applying Neural Networks to Pharmacokinetic Data Sets," *Decision Support Systems* (30:2), 2000, pp.139-152.

Vapnik, V., *The Nature of Statistical Learning Theory*, Springer-Verlag, 1995.

Vincent, J. L., Bihari, D. J., Suter, P. M., Bruining, H. A., White, J., Nicolas-Chanoin, M. H., Wolff, M., Spencer, R. C., and Hemmer, M., "The Prevalence of Nosocomial Infection in Intensive Care Units in Europe: Results of the European Prevalence of Infection in Intensive Care (EPIC) Study. EPIC International Advisory Committee," *The Journal of the American Medical Association* (274:8), 1995, pp.639-644.

Wang, Y. and Witten, I. H., "Induction of Model Trees for Predicting Continuous Classes," *Proceedings of the European Conference on Machine Learning*, Prague, Czech Republic, 1997, pp.128-137.

Winter, M. E., *Basic Clinical Pharmacokinetics*, 4th Ed., Lippincott Williams and Wilkins, 2003, p.451.

Yamamura, S., Takehira, R., Kawada, K., Nishizawa, K., Katayama, S., Hirano, M., and Momose, Y., "Application of Artificial Neural Network Modelling to Identify Severely Ill Patients whose Aminoglycoside Concentrations Are Likely to Fall Below Therapeutic Concentrations," *Journal of Clinical Pharmacy and Therapeutics* (28), 2003, pp.425-432.

Zhou, Z. H., Wu, J., and Tang, W., "Ensembling Neural Networks: Many Could Be Better Than All," *Artificial Intelligence* (137:1-2), 2002, pp.239-263.