Paradigms, benefits,
and challenges

Drug Dosing Control in Clinical Pharmacology

By James M. Bailey
and Wassim M. Haddad

Control technology influences modern medicine through robotic surgery, electrophysiological systems (pacemakers and automatic implantable defibrillators), life support (ventilators and artificial hearts), and image-guided therapy and surgery. An additional area of medicine suited for applications of control is clinical pharmacology, in which mathematical modeling plays a prominent role [1], [2]. Although numerous drugs are available for treating disease, proper dosing is often imprecise, resulting in increased costs, morbidity, and mortality. In this article, we discuss potential applications of control technology to clinical pharmacology, specifically the control of drug dosing [3]. (See "Control Engineering and Medicine: A Fruitful Collaboration.")

We begin by considering how dosage guidelines are developed. Drug development begins with animal experimentation. Promising agents are moved to human trials, which begin with healthy volunteers and progress to patients with the specific disease for which the drug is being developed. Early stages of these trials focus on safety, while the final trials usually involve administration of a placebo and different drug doses to evaluate efficacy. Efficacy is defined statistically, and aggregate therapeutic effects do not preclude the existence of individual patients for whom the drug is either not efficacious or causes side effects. If a therapeutic effect is observed, then the drug may be approved by the Food and Drug Administration (FDA). In general, the recommended dosage is the level found to be efficacious in the "average" patient. Herein lies the problem: the "average" patient does not exist.
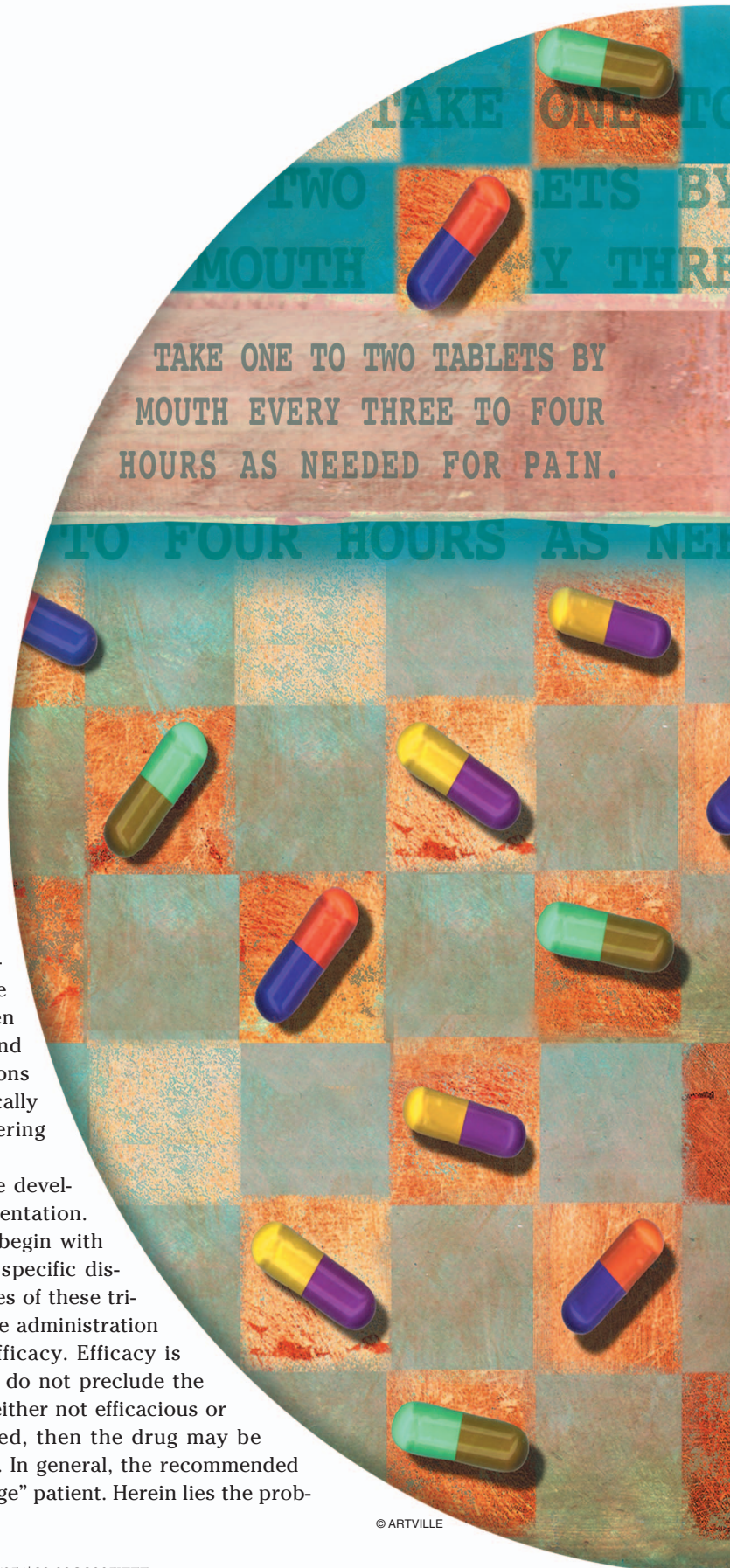
© ARTVILLE

Control Engineering and Medicine: A Fruitful Collaboration

The barriers between control engineering and medicine are slowly eroding as it becomes more evident that control system technology has a great deal to offer medicine. Our interdisciplinary collaboration is based on the desire of one author to explore the area of biomedical control engineering and the interest of the other author in applications of pharmacokinetic and pharmacodynamic modeling. Our collaboration began with e-mail communication, followed by a period of mutual education. It then progressed to monthly meetings to discuss and identify control methodologies and paradigms for addressing pharmacological problems. From our own experience, it takes somewhat longer for the clinician to become comfortable with the vocabulary and concepts of control than vice versa. While anesthesiologists use feedback control every working day, few are familiar with the mathematical rigor of the control scientist.

Our collaboration is facilitated by the fact that while pharmacokinetics, the clinician's research interest, is based on dynamical systems theory, the level of rigor in pharmacokinetics is well below that of control science. Thus, the rate-limiting step in the collaboration is introducing the clinician to the vocabulary of the control engineer. The collaboration is now at a point where the clinician can serve on the dissertation committee of the control engineer's graduate students. Furthermore, the clinician serves a valuable role in the collaboration by giving the control engineer an idea of which assumptions and approximations are clinically realistic. This exchange is facilitated by having the control engineer observe the clinician in the operating room and the intensive care unit, as we have done. This interaction is particularly important as theory transitions to clinical implementation.

Substantial variability exists among patients both in the drug concentration at the locus of the effect (the *effect-site concentration*) resulting from a given dose and in the therapeutic efficacy of a given effect site concentration. Frequently, the appropriate dose for a specific patient is found by trial and error. For example, a physician treating a patient for hypertension typically begins by prescribing the recommended dose. After observing the effect of the drug on blood pressure, the doctor adjusts the dose empirically.

A Primer on Clinical Pharmacology

Pharmacokinetic Models

Drug dosing can be made more precise by using *pharmacokinetic* and *pharmacodynamic* modeling [4]. Pharmacokinetics is the study of the concentration of drugs in tissue as a function of time and dose schedule. Pharmacodynamics is the study of the relationship between drug concentration and drug effect. By relating dose to resultant drug concentration (pharmacokinetics) and concentration to effect (pharmacodynamics), a model for drug dosing can be generated.

The distribution of drugs in the body depends on transport and metabolic processes, many of which are poorly understood [1], [4]. However, *compartmental models*, that is, dynamical models based on conservation laws that capture the exchange of material between coupled macroscopic subsystems or compartments, are widely used to model these processes [2]. Pharmacokinetic compartmental models typically assume that the body is comprised of more than one compartment. Within each compartment, the drug concentration is assumed to be uniform due to perfect, instantaneous mixing. Transport

Pharmacokinetic Terminology

Pharmacokinetic models are frequently described in terms of half lives, compartment volumes, and clearance. For instance, most pharmacokinetic papers report the *terminal elimination half life*, which is the time required for drug concentration to decrease by 50% if all tissues were equilibrated with the blood concentration. This parameter is useful to clinicians desiring a measure of the duration of drug effect once the drug is discontinued from chronic use. Similarly, pharmacokinetic papers sometimes report *distribution half lives*, the time needed for 50% equilibration between two compartments when a drug is initially confined to one compartment. Another parameter is *clearance*, a term originally taken from the renal physiology literature. Clearance is defined as the effective volume of tissue or blood cleared of drug per unit time. Elimination clearance is equal to the product of the compartment volume and the elimination rate from the compartment. While drugs are not eliminated in the simplistic fashion implied by the above definition, the concept is useful since the constant drug input rate needed to achieve a given drug concentration is equal the product of the desired concentration and clearance.

to other compartments and elimination from the body occur through metabolic processes. For simplicity, the transport rate is often assumed to be proportional to drug concentration. Although the assumption of instantaneous mixing is an idealization, it has little effect on the accuracy of the model as long as we do not try to predict drug concentrations immediately after the initial drug dose.

In a simple one-compartment model, the body is assumed to consist of a single compartment in which instantaneous mixing occurs, followed by elimination. It is usually assumed that elimination is linear, with the rate of elimination directly proportional to the drug concentration in the compartment. This model is characterized by two parameters: the compartmental volume V_d and the elimination rate constant a_e. For this simple model, the concentration C (in moles/volume or mass/volume) immediately after a dose of mass D is equal to D/V_d, and the drug is subsequently eliminated at the rate a_eC with exponential decay. While the behavior of a few drugs can be described adequately using this model, the model is too simplistic for most drugs. Furthermore, for drugs taken orally, a model with two or more compartments is required. One compartment represents the gastro-intestinal tract, which receives the dose and transfers it to a second compartment. The second compartment represents *intravascular blood* (blood within arteries or veins) and other organ systems.

A two-compartment *mammillary* model [2] can also be used for drugs administered intravenously. This model includes a central compartment that receives the intravenous dose with instantaneous mixing and is typically identified with organs such as the heart, brain, liver, and kidney. These organ systems receive a large amount of blood flow per unit mass and, hence, are well mixed with intravascular blood. The drug is then transferred to a peripheral compartment comprised of muscle and fat or it is metabolized and eliminated from the body. For most drugs, the enzyme systems responsible for drug metabolism are found in the liver or kidney so that metabolism in the peripheral compartment can be ignored. Drug in the peripheral compartment transfers back to the central compartment with linear kinetics.

State-Space Models

The two-compartment mammillary system is described by the state-space model

$$\dot{x}(t) = Ax(t), \quad x(0) = x_0, \quad t \geq 0, \tag{1}$$

where

$$A = \begin{bmatrix} -(a_{21} + a_{11}) & a_{12} \\ a_{21} & -a_{12} \end{bmatrix},$$

$x = [x_1, x_2]^{\mathrm{T}}$ is the state vector representing the masses in the two compartments; a_{12} and a_{21} are the nonnegative
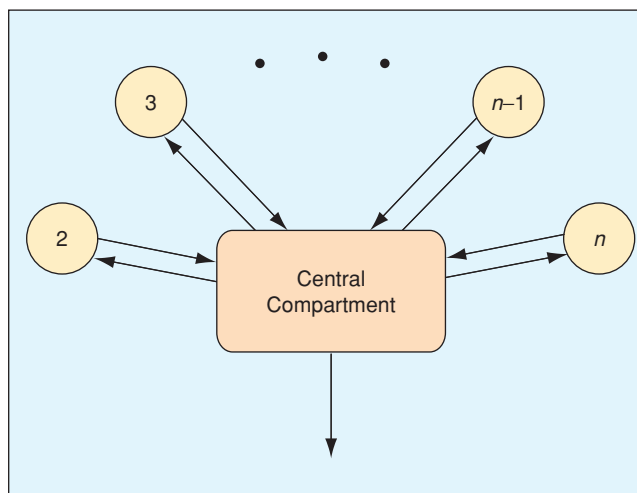

Figure 1. *The n-compartment mammillary model. The central compartment, which is the site for drug administration, is generally thought to be comprised of the intravascular blood volume as well as highly perfused organs such as the heart, brain, kidney, and liver. The central compartment exchanges the drug with the peripheral compartments comprised of muscle, fat, and other organs and tissues of the body, which are metabolically inert as far as the drug is concerned.*

transfer coefficients from compartment 2 to compartment 1 and from compartment 1 to compartment 2, respectively; and the nonnegative coefficient a_{11} is the rate at which drug is eliminated from the system through the central compartment 1. Entry $(2, 2)$ of A reflects no elimination from compartment 2. An additional parameter is the volume V_1 of the central compartment, for a total of four pharmacokinetic parameters. (See "Pharmacokinetic Terminology.") Note that with the assumption of instantaneous mixing, the concentration at $t = 0^+$ after dose D is administered is D/V_1.

The two-compartment mammillary model is useful for many drugs administered intravenously. The basic two-compartment model predicts that the drug concentration in the central compartment after a bolus (impulsive) initial dose can be described by the sum of two terms decreasing exponentially with time. However, to fit the data, some drugs require three or more exponential terms, which motivates an extension of the two-compartment model [2]. Figure 1 shows an *n*-compartment mammillary model, which includes a central compartment that distributes drug into the interstitial spaces of several organs and tissues of the body.

In most cases, the assumption of linear transfer is maintained so that the system is modeled by (1), where $x \in \mathbb{R}^n$ represents the system compartmental masses or system compartmental concentrations and $A \in \mathbb{R}^{n \times n}$ is a *compartmental matrix* [2] when x represents compartmental masses and a *Metzler matrix* [2] when x represents

compartmental concentrations. Hence, (1) describes a nonnegative, compartmental dynamical system [2].

Compartmental pharmacokinetic models, especially mammillary models, are coarse-grained oversimplifications. Consider the injection of a drug into a small peripheral vein in the hand. The drug is transported by the venous stream to the right atrium and the right ventricle, binding to blood cells or proteins and mixing with venous streams as veins coalesce. Large-scale mixing occurs in the right atrium and ventricle, transporting the drug to the lung, where some of the drug can bind to tissue. From the lung, the drug returns to the left atrium and ventricle. There, it is expelled into the aorta for transport to other inert tissues, where drug binding occurs, and to the liver and kidney, where the drug is metabolized. Modeling this process with a small number of compartments is thus an approximation.

Drug Action, Effect, and Interaction

The clinical utility of pharmacokinetic models depends entirely on the time scale of the application. For example, these models work well for determining suitable dosing intervals for drugs administered orally as well as for maintaining appropriate anesthetic concentrations during surgery. Using simplified mammillary models, it is possible to achieve median absolute performance errors (the normalized difference between target and measured anesthetic concentrations) of less than 20% when drug concentrations are sampled every 15 minutes. This level of performance is clinically acceptable since drug concentrations within this range generally achieve the desired effect.

We now consider the problem of predicting drug concentrations during the administration of anesthesia. Anesthesia is typically initiated by administering a bolus dose of a hypnotic drug intravenously. During the first few minutes after initiation of intravenous anesthesia by administration of a bolus dose, mammillary compartmental models fail to accurately predict drug concentrations because of the assumption of instantaneous mixing. More elaborate models postulate multiple compartments in series to approximate the transport of the drug from the site of injection to the central circulation. Additional parallel compartments (similar to mammillary models) account for drug distribution to peripheral tissue (muscle and fat). These extensions help to describe drug concentration immediately after a bolus dose initiation of anesthesia [5].

While the most commonly used pharmacokinetic models are linear, the underlying processes that determine pharmacokinetic behavior are nonlinear. For example, the molecular processes of drug metabolism are described by Michaelis-Menten kinetics, in which the rate of drug metabolism is given by $V_m C / (K_m + C)$, where V_m is the maximum rate of reaction, K_m is the drug concentration that achieves 50% of the maximal effect, and C is the drug concentration. However, large-scale pharmacokinetic models assume linear drug metabolism or elimination. Similarly, most compartmental models assume that drug distribution between tissues is linear. However, delivery of drug to tissue per unit time is equal to the product of drug concentration in the blood and *regional perfusion*, the blood flow to tissue per unit time. Regional perfusion may be nonlinearly related to the drug concentration since anesthetic drugs alter cardiovascular function.

Pharmacokinetic Parameter Estimation

Data used for pharmacokinetic modeling is collected by administering a drug to patients, drawing blood samples at designated times after the initiation of dosing, and determining the concentration of the drug as a function of time (but not in real time). Consequently, most pharmacokinetic investigations focus on blood concentrations. One goal of this analysis is to derive an expression for the *unit disposition function*, the time-dependent blood concentration that results from a unit bolus dose. Assuming linear kinetics, if the unit disposition function f_{ud} is known, then the resulting blood concentration is given by the convolution integral

$$C(t) = \int_0^t f_{ud}(\tau) D(t - \tau) d\tau, \qquad (2)$$

where $D(t)$ is the dose as a function of time [6].

It is not feasible to measure drug concentration in the tissue at the site of the therapeutic effect. Since drugs are distributed to the action site by blood flow, the effect site rapidly equilibrates with blood, and it is often assumed that effect-site concentration and blood concentration are equal. If the equilibration time between the central intravascular blood volume and the effect site is clinically relevant, then the pharmacokinetic model must be revised to include an effect-site compartment distinct from the central compartment.

Pharmacokinetic model parameters that comprise the system matrix A are estimated by fitting models to the data. There are numerous sources of noise in the data, from assay error to human recording error. Because of model approximation and noise, there is always an offset between the concentration predicted by the model and the observed data, the prediction error. One method for estimating pharmacokinetic parameters is maximum likelihood [7]. This approach assumes a statistical distribution for the prediction error and then determines the parameter values that maximize the likelihood of the observed results. Suppose we conduct a study in a single patient from whom we collect blood samples at ten different points in time after a single intravenous bolus. If we assume that the prediction error for an individual patient

(*intrapatient error model*) has a normal or Gaussian distribution, then the likelihood $\mathcal{L}$ of the observed results is given by

$$\mathcal{L} = \prod_{i=1}^{r} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-PE_i^2/2\sigma^2}, \tag{3}$$

with the prediction error PE_i of the ith observation given by $PE_i = C_{p_i} - C_{m_i}$, where C_{p_i} is the ith predicted drug concentration and C_{m_i} is the ith measured drug concentration, σ^2 is the variance of the Gaussian distribution of prediction errors, and r is the number of observations (measured concentrations).

The likelihood (3) of the observed results is a function of σ and the pharmacokinetic parameters. By maximizing (3) (or, more commonly, its logarithm) with respect to the pharmacokinetic parameters and σ, one can estimate the structural model parameters (the entries of the system matrix A) and the error model parameters (in this case, σ) that maximize the likelihood of the observed results. The above example reduces to least squares estimation when σ^2 is a constant. Using a more sophisticated error model, in which σ^2 is proportional to a power of the predicted concentration, leads to weighted least squares estimation [7].

There are two distinct approaches to estimating mean pharmacokinetic parameters for a population of patients [8], [9]. In the first approach, models are fitted to data from individual patients and the pharmacokinetic parameters are then averaged (*two-stage analysis*) to provide a measure of the pharmacokinetic parameters for the population. The second approach, called *mixed-effects modeling*, is to pool the data from individual patients. In this situation, the prediction error is determined by the stochastic noise of the experiment as well as by the fact that *different patients have different pharmacokinetic parameters*. The statistical model used to account for the discrepancy between observed and predicted concentrations must take into consideration not only variability between observed and predicted concentrations within the same patient (*intrapatient* variability), but also variability between patients (*interpatient* variability). Most commonly, it is assumed that the interpatient variability of pharmacokinetic parameters conforms to a log-normal distribution. This sophisticated method of analysis estimates the mean structural pharmacokinetic parameters as well as the statistical variability of these elements in the population.

Analysis based on mixed-effects modeling is powerful for two reasons. First, this approach gives the clinician an estimate of both the pharmacokinetic parameters and their variance. These statistics are important for the clinician since no matter how desirable the properties of a drug are on average, extreme variability in these parameters may indicate that the drug is not safe for clinical use. Second,

mixed-effects modeling may allow a reduction in the amount of data needed from each patient. In a two-stage analysis, one must have enough data points from each patient to estimate the patient's pharmacokinetic parameters. For example, a two-compartment mammillary model requires four pharmacokinetic parameters. With mixed-effects modeling, it is possible to estimate these parameters for any one patient with four or fewer data points.

Pharmacodynamic Models

In contrast to pharmacokinetic modeling, pharmacodynamic modeling is less readily related to molecular processes. The molecular mechanism of action of many drugs is well understood; most drugs act by binding to a receptor on or within target cells [4]. There is a well-developed theory of multiple equilibrium binding of ligands, such as drug molecules, to receptors on larger macromolecules, such as proteins. In theory, pharmacodynamics, which models the relationship between drug concentration and effect, should follow from models of molecular binding. However, the physiological effect is an interplay of numerous factors, and it is generally not possible to analytically relate the drug effect at the level of the intact organism to the number of receptors bound by the drug at the molecular level. Empirical models are thus needed. It might be assumed that drug effect is proportional to the drug concentration at the effect site, but this simple linear model is unrealistic since it admits the possibility of limitless drug effect as drug concentration increases while ignoring saturation effects.

One empirical pharmacodynamic model is given by the *Hill equation*

$$E = E_{\max}C^{\gamma} / \left(C^{\gamma} + C_{50}^{\gamma}\right), \tag{4}$$

where E is the drug effect, $E_{\max}$ is the maximum drug effect, C is the drug concentration, C_{50} is the drug concentration associated with 50% of the maximum effect, and γ is a dimensionless parameter that determines the steepness of the concentration-effect relationship [10]. This model was developed in 1906 to describe a *molecular interaction*, the binding of oxygen to hemoglobin. Since that time the Hill model has been applied to various phenomena that are far removed from explanations at the molecular level. A number of modified versions of this model have been employed, including the case in which the drug effect is a binary (yes-or-no) variable. An example of a binary variable is anesthesia, for which the patient is either responsive or not. In this case, $E_{\max} = 1$, and the pharmacodynamic model becomes

$$P = C^{\gamma} / \left(C^{\gamma} + C_{50}^{\gamma}\right),$$

where the effect is now the probability P that the patient does not respond to some noxious stimuli [11], [12].

In typical pharmacodynamic studies, a drug is administered and the effect is measured by taking a blood sample at various points in time to determine the drug concentration at the time of observation of effect. The parameters of the pharmacodynamic model (E_{max}, C_{50}, γ) can then be estimated by the maximum likelihood or generalized least squares methods described above. Obviously, if drug concentrations in the blood and effect site have not equilibrated, this analysis does not apply.

It should be noted that pharmacodynamic models are inherently nonlinear. This property is in contrast to the linearity of pharmacokinetic models. However, the interplay with pharmacodynamics may also lead to nonlinear pharmacokinetics. For example, some intravenous anesthetics depress *cardiac output*, the volume of blood pumped by the heart per unit time. Since the basic transport processes that determine pharmacokinetic behavior depend on blood flow, administration of the drug alters its kinetics. Furthermore, since the pharmacodynamic relationship between drug concentration and depression of cardiac output is nonlinear, the pharmacokinetics of the drug are, in reality, also nonlinear.

Clinical Pharmacology and Drug Dosing

Open-Loop Drug Dosing
In addition to safety and efficacy, the FDA requires that, prior to the approval of any new drug, manufacturers provide a pharmacokinetic evaluation to establish a basis for dosage guidelines. The concentration that results from drug administration is determined by the transport processes that distribute the drugs to various tissues as well as by the metabolic processes that transform the drug. However, the vast majority of drugs are given for chronic conditions; when the time scale of treatment greatly exceeds the time scale of the distributive processes, there will be equilibration of drug through the various tissues.

The pharmacokinetics of most drugs given chronically are described by (1), where A is a scalar. Drug calculations are now greatly simplified. For example, consider an antihypertensive drug with a *half life* (the time needed for the drug concentration to decrease by 50% after discontinuation of administration) of 12 hours. If a dose of 50 mg is efficacious in the average patient, then a suitable dosing schedule would be an initial dose of 50 mg with subsequent dosing of 25 mg every 12 hours. As another example, suppose that a blood concentration of an intravenous anesthetic of 100 μg/ml reliably produces unconsciousness and the *clearance* (the effective volume of blood cleared of drug per unit time) is 150 ml/minute. An infusion rate of 100 μg/ml $\times$ 150 ml/min $=$ 15000 μg/min maintains the desired blood concentration, although this concentra-

tion is achieved only when distributive processes have equilibrated. Many of the dosing guidelines recommended by the manufacturers of drugs are based on simple calculations like these.

There have been attempts to develop more precise open-loop control in acute care, especially in anesthetic pharmacology. Since the appearance of computers in the operating room in the 1980s, investigators have developed computer-controlled pump systems that continually adjust the drug infusion rate to achieve and maintain the drug concentration desired by the clinician [13]–[16]. These systems use the pharmacokinetic model

$$\dot{x}(t) = Ax(t) + Bu(t), \quad x(0) = x_0, \quad t \geq 0$$

to calculate the dose $u(t)$ needed to achieve and maintain the target drug concentration. To implement this approach, it is necessary to know the pharmacokinetic parameters that define A and B. In open-loop control, specific pharmacokinetic parameters are not known for the individual patient. Instead, it is assumed that *average* parameter values, taken from the pharmacokinetic literature, are applicable to the individual patient. The dose calculated using these average parameter values is then delivered by a pump controlled by the computer. Despite the obvious fact that these systems ignore interpatient pharmacokinetic variability, studies have demonstrated that drug concentrations are better maintained in therapeutic ranges by open-loop control than with standard clinical practice.

Closed-Loop Drug Dosing
While initial dosing guidelines are often based on the average patient, the significant interpatient pharmacokinetic and pharmacodynamic variability observed for most drugs suggests that *precise* drug dosing requires closed-loop control. Most patients, especially those treated for chronic disease, are familiar with this closed-loop process. The physician prescribes an initial dose, observes the response, and adjusts the dose. Although some physicians are adept at this process, it is usually time consuming. Also, the efficiency of the process depends on the experience of the clinician since there is not enough data available on many drugs to develop quantitative guidelines.

The process of dose titration can be made more precise by using mixed-effects pharmacokinetic modeling and post-hoc Bayesian estimation of individual patient pharmacokinetic parameters [7]–[8]. Recall that mixed-effects modeling provides not only estimates of pharmacokinetic parameters, but also their variance within the population. Suppose one or more drug concentrations are measured in an individual patient. Using Bayesian probability principles, the likelihood of a given value of a pharmacokinetic

parameter Θ is proportional to $P(C|\Theta)P(\Theta)$, where $P(C|\Theta)$ is the probability of the observed concentration C as a function of Θ. Furthermore, $P(\Theta)$ is the a priori probability of a given value of Θ, which is given by the assumed distribution of Θ (as noted above, usually lognormal) with the variance of Θ estimated from the mixed-effects analysis.

By determining the mode of $P(C|\Theta)P(\Theta)$ with respect to Θ (the value of Θ at which $P(C|\Theta)P(\Theta)$ is maximized), one can derive a maximum likelihood estimate of Θ for the specific patient. The patient-specific parameter estimate can be used to calculate the dose needed to achieve a given drug concentration [17]. However, because of pharmacodynamic variability, more precise control of drug *concentration* does not necessarily lead to better control of drug *effect*.

The process of titrating drug dose to the desired effect may be acceptable for chronic outpatient therapy; yet, in the acute care environment, such as the operating room or the intensive care unit, this process is often either dangerously slow or imprecise. Feedback control of drug effect, in contrast with drug concentration, has much to offer modern medicine. The remainder of this article focuses on drugs used in the acute care setting.

To implement closed-loop control in an acute care environment, real-time measurement of drug effect is required. Early attempts at closed-loop control focused on regulating variables that are conveniently measured. By their very nature, cardiovascular and central nervous system functions are critical in the acute care environment. Thus, technologies have evolved for their measurement. The primary applications of closed-loop drug administration are hemodynamic management and control of consciousness. Next, we review closed-loop control of the cardiovascular system, which illustrates problems inherent in the application of control technology to physiological function.

Closed-Loop Control of Cardiovascular Function

A major side effect of cardiac surgery is that patients can become hypertensive [18], requiring treatment to prevent cardiac dysfunction, pulmonary edema, myocardial ischemia, stroke, and bleeding from fragile sutures. Although drugs are available for treating post-operative hypertension, titrating these drugs to regulate blood pressure is often difficult. Underdosing leaves the patient hypertensive, whereas overdosing can reduce the blood pressure to levels associated with shock. Since the late 1970s, there has been interest in developing controllers for administering sodium nitroprusside (SNP), a commonly used and potent antihypertensive.

The problems encountered in hemodynamic control are enlightening. Initial attempts used simple nonadaptive methods such as proportional-derivative (PD) or proportional-integral-derivative (PID) controllers, which assume a linear relationship between infusion rate and effect [19], [20]. However, while the drug concentration is given by the convolution of the infusion rate and a transfer function as in (2), the relationship between effect and infusion rate is nonlinear as in (4). Also, a significant challenge to the design of a blood-pressure controller is the time delay between drug administration and the clinical effect, which can lead to system oscillations. Although early blood pressure controllers included time delays in the system model, the delays were assumed to be the same for each patient [19], [20].

While early controllers were successful in some patients, these techniques have not been widely implemented due to nonlinear patient response and differences in drug sensitivity among patients. Interpatient variability, as well as a patient's sensitivity to drugs, motivated the development of single-model and multiple-model adaptive controllers [21], [22]. Single-model adaptive controllers are based on on-line estimation of system parameters using minimum variance or least squares methodologies. These controllers perform poorly due to large amplitude transients [21].

Multiple-model adaptive controllers represent the system by means of a finite number of models. For each model, there is a separate controller. The probability that the system is represented by each of the different models is calculated from the relative offsets of the system response and the response predicted by each model. The output of the controller is the probability-weighted sum of the outputs from each model [23], [24]. Multiple-model adaptive controllers have proven to be more satisfactory than single-model adaptive controllers and fixed-gain controllers [24]. Subsequent refinements to blood pressure control include both single-model, reference adaptive control [25], which appears promising in simulations, and neural network-based methods [26]. There is also interest in optimal control since SNP has toxic side effects when the dose is too high [27].

These investigations into controlling blood pressure reveal the challenges inherent to biological systems: nonlinearity, interpatient variability (system uncertainty), and time delay. Despite the refinements of closed-loop blood pressure controllers, such controllers are seldom used clinically. Although blood pressure control is important, cardiovascular function involves several other important variables, all of which are interrelated [18]. The intensive care unit clinician must ensure not only that blood pressure is within appropriate limits, but also that cardiac output (the amount of blood pumped by the heart per minute) is acceptable and that the heart rate is within reasonable limits. Mean arterial blood pressure is proportional to cardiac output, with the proportionality constant denoting the systemic vascular resistance, in analogy with

Ohm's law. Cardiac output is equal to the product of heart rate and *stroke volume*, the volume of blood pumped with each beat of the heart. Stroke volume, in turn, is a function of *contractility* (the intrinsic strength of the cardiac contraction), *preload* (the volume of blood in the heart at the beginning of the contraction), and *afterload* (the impedance to ejection by the heart).

The intensive care unit clinician must balance all of these variables. Inotropic agent drugs, or drugs that increase the strength of contraction of the heart, also have variable effects on heart rate and afterload. There are also *vasopressor* drugs, which increase afterload, and *vasodilator* drugs, which decrease afterload. Finally, stroke volume can be improved by giving the patient intravenous fluids and increasing preload. However, too much fluid can potentially be deleterious by impairing pulmonary function as fluid builds up in the lungs. The fact that closed-loop control of blood pressure has not been widely adopted by clinicians is not surprising when one considers the complex interrelationships among hemodynamic variables. Future applications of control theory in the form of adaptive and robust optimal controllers that oversee the administration of multiple drugs (inotropes, vasopressors, and vasodilators) and fluids will be a major advance in critical care medicine. Preliminary investigation of the control of multiple hemodynamic drugs has already begun [28], [29].

Closed-Loop Control of Anesthesia

Automated Anesthesia: Inhalational and Intravenous

Anesthesia involves several components: *analgesia*, lack of reflex response (such as increased blood pressure or heart rate) to surgical stimulus; *areflexia*, lack of movement (which simplifies the task of the surgeon); and *hypnosis*, or lack of consciousness. Closed-loop control of anesthesia may be implemented either by controlling these components of anesthesia or, more simply, by controlling the anesthetic concentration and assuming that the appropriate concentration will lead to the desired effect.

Real-time spectroscopic methods for measuring the concentration of inhaled anesthetic agent in *end-tidal* (that is, exhaled) gases are now routinely available in most operating rooms. End-tidal anesthetic gas concentration is a reasonable surrogate for arterial blood anesthetic concentration [30]. End-tidal anesthetic agent concentrations can be measured in real time, which facilitates closed-loop control of end-tidal anesthetic concentration. However, anesthetic concentration cannot be equated with anesthetic effect. More recently, real-time processed *electroencephalograph* (EEG) measurement has offered the possibility of closed-loop control of anesthetic effect. It has been known for decades that the induction of anesthesia causes changes in the EEG [31]. In the last decade, there has been substantial progress in developing processed EEG monitors that measure the depth of anesthesia to provide performance variables for closed-loop controllers [31].

Inhaled anesthetic agents have been the mainstay of clinical practice since the first delivery of anesthesia. A fundamental characteristic of every inhaled anesthetic agent is its "MAC" value, *m*inimum (in suppressing the response to painful stimuli) *a*lveolar (alveoli are the functional units of the lung) *c*oncentration, which is associated with a 50% probability of patient movement in response to surgical stimulus [28]. By maintaining end-tidal concentrations well above MAC, the practitioner is relatively assured of hypnosis. The ready availability of spectroscopic systems for measuring end-tidal anesthetic concentration in real time has led several investigators to develop closed-loop controllers.

The earliest anesthesia controllers use fixed-gain PID controllers [32], [33], which assume that all patients are the same. In contrast, the adaptive model-based controllers in [34] and [35] rely on least squares methods to estimate the patient-specific system parameters. In animal studies [23], the adaptive controllers have performed more effectively than the fixed-gain controllers. However, the adaptive controllers have not been widely adopted clinically since control of anesthetic concentration does not translate into control of anesthetic effect due to interpatient pharmacodynamic variability.

EEG-Based Control

Closed-loop control of anesthesia requires a monitor of anesthetic effect, specifically consciousness, which has been an elusive challenge for anesthesiologists. The EEG, which measures electrical activity in the brain, has been an obvious candidate. In particular, neurophysiologists have observed that the EEG of an anesthetized patient contains slower waves with higher amplitudes. However, the EEG is comprised of multiple time series and multiple spectra and, while anesthesia induces characteristic changes in the EEG, it is not clear which, if any, characteristic of the EEG best reflects the anesthetic state.

EEG-based closed-loop control of anesthesia was first proposed in [36]. Subsequently, a closed-loop, model-based adaptive controller was developed and clinically tested in [37], delivering intravenous anesthesia using the median frequency of the EEG power spectrum as the regulated variable. The model used in [37] assumes a two-compartment pharmacokinetic model for which the drug concentration $C(t)$ after a single bolus dose is given by

$$C(t) = Ae^{-\alpha t} + Be^{-\beta t}, \quad t \geq 0,$$

where A, B, α, and β are patient-specific pharmacokinetic parameters. It is also assumed that the median EEG frequency E is related to the drug concentration by the modified Hill equation

$$E = E_0 - E_{\max} \left[C^\gamma / \left(C^\gamma + C_{50}^\gamma \right) \right], \qquad (5)$$

where E_0 is the baseline signal, $E_{\max}$ is the maximum decrease in signal with increasing drug concentration, C_{50} is the drug concentration associated with 50% of the maximum effect, and the parameter γ describes the steepness of the concentration-effect curve. From (5) it can be seen that the drug effect is a function of the pharmacokinetic parameters A, B, α, and β as well as the pharmacodynamic parameters E_0, $E_{\max}$, C_{50}, and γ. If these parameters are known, it is straightforward to calculate the dose regimen needed to achieve the target EEG signal. However, these parameters are not known for individual patients, and interpatient variability may be significant. Estimates for the coefficients of variability for some parameters are as high as 100%.

The algorithm in [37] assumes that the pharmacodynamic parameters E_0, $E_{\max}$, C_{50}, and γ as well as the pharmacokinetic parameters α and β are equal to the mean values reported in prior studies. Using the mean values of the pharmacokinetic parameters A and B from prior studies as starting values, estimates of these parameters are refined by analyzing the difference ΔE between the target and observed EEG signal. Linearizing ΔE with respect to A and B yields

$$\Delta E = (\partial E / \partial A)\delta A + (\partial E / \partial B)\delta B, \qquad (6)$$

where δA and δB represent the updates to the values of A and B in the adaptive control algorithm. In conjunction with minimizing $(\delta A)^2 + (\delta B)^2$, (6) is used to estimate δA and δB. This algorithm is only partially adaptive in that A and B are the only parameters of the model that are updated. This algorithm was implemented for the intravenous anesthetic agents methohexital and propofol but did not appear to offer great advantage over standard manual control [37], [38]. The observed performance might have been due to the approximations of the algorithm or the deficiencies of the median EEG frequency as a measure of the depth of anesthesia.

Bispectral Index-Based Control

Since the work of [37], alternative EEG measures of depth of anesthesia have been developed. Possibly the most notable of these measures is the *bispectral index* or BIS [39], [40]. The BIS is a single-composite EEG measure, which appears to be closely related to the level of consciousness (see Figure 2). The BIS signal is related to drug concentration by the empirical relationship

$$\mathrm{BIS}(c_{\mathrm{eff}}) = \mathrm{BIS}_0 \left(1 - \frac{c_{\mathrm{eff}}^\gamma}{c_{\mathrm{eff}}^\gamma + \mathrm{EC}_{50}^\gamma} \right), \qquad (7)$$

Figure 2. *Bispectral index (BIS) monitor. A sensor is placed on the patient's forehead to measure electrical activity in the brain and monitor the patient's level of consciousness.*

where BIS_0 denotes the baseline (awake state) value, which, by convention, is typically assigned a value of 100; c_{eff} is the drug concentration in μg/ml in the effect-site compartment (brain); EC_{50} is the concentration at half maximal effect and represents the patient's sensitivity to the drug; and γ determines the degree of nonlinearity in (7). Here, the effect-site compartment is introduced to account for finite equilibration time between the central compartment concentration and the central nervous system concentration [41].

In [42], closed-loop control is used to deliver the intravenous anesthetic propofol based on a model-based adaptive algorithm with the BIS as the regulated variable. The algorithm in [42] is based on a pharmacokinetic model that predicts the drug concentration as a function of infusion rate and time and uses a pharmacodynamic model that relates the BIS signal to concentration.

In contrast to [37], it is assumed in [42] that the pharmacokinetic parameters are always correct and that differences in individual patient response are due to pharmacodynamic variability. Moreover, the approach of [42] predicts the anesthetic concentration using the pharmacokinetic model and then constructs a BIS-concentration curve using both the observed BIS during induction and the predicted propofol concentration. During each time epoch, the difference between the target BIS signal and the observed BIS signal is used to update the pharmacodynamic parameters relating concentration and BIS signal for the individual patient. However, this algorithm does not update the pharmacokinetic parameters.

The results in [42] demonstrate excellent performance as measured by the difference between the target and observed BIS signals. However, as pointed out in [43], the

performance of the model-based adaptive controller may reflect the fact that the patient was not fully stressed. In [42], a high dose of the opioid remifentanil, a neurotransmitter inhibitor resulting in significant analgesic effect, was administered in conjunction with propofol. Consequently, central nervous system excitation due to surgical stimulus was blunted and, thus, the need to adjust the propofol dose as surgical stimulus varied was diminished. It is unknown whether the control system would have been effective in the absence of deep narcotization.

In contrast to the model-based adaptive controllers in [37], [38], and [42], a PID controller using the BIS signal as the variable to control the infusion of propofol is considered in [44]. The median absolute performance error, the median value of the absolute value of $\Delta E/E_{\text{target}}$, was good (8.0%), although in three of the ten patients, oscillations of the BIS signal around the setpoint were observed, and anesthesia was deemed clinically inadequate in one of the ten patients. The same system was used in [45], with an auditory-evoked potential based on somatosensory information provided by auditory stimulation generating oscillations within the EEG signal as the regulated performance variable.

Intravenous propofol anesthesia was delivered in [44] by means of a fuzzy logic closed-loop controller that uses both auditory (constant frequency, constant amplitude signal delivered by earphones) evoked responses and cardiovascular responses as the regulated variables. This system has had only minimal clinical testing [46]. More recently, [47] considers model-based controllers for inhalation anesthetic agents that attempt to control the BIS signal or mean arterial blood pressure, while keeping end-tidal anesthetic concentrations within prespecified limits.

Nonlinear Adaptive and Neuro Adaptive Control for General Anesthesia

Because of the uncertainties in the pharmacokinetic and pharmacodynamic parameters due to interpatient variability, we have developed adaptive controllers that can be implemented using the processed EEG as a performance variable (see Figure 3). Using compartmental

models, a Lyapunov-based direct adaptive control framework developed in [48] and [49] guarantees partial asymptotic setpoint stability of the closed-loop system (asymptotic setpoint stability with respect to part of the closed-loop system states associated with the physiological state variables). Furthermore, the remaining states associated with the adaptive controller gains are bounded. The adaptive controllers, which are constructed without requiring knowledge of the system pharmacokinetic and pharmacodynamic parameters, provide a nonnegative control input for stabilization with respect to a given setpoint in the nonnegative orthant.

In [50] and [51], we present a neural network adaptive control framework that accounts for combined interpatient pharmacokinetic and pharmacodynamic variability. In particular, we develop a framework for adaptive setpoint regulation of nonlinear uncertain compartmental systems. The formulation in [50] and [51] addresses adaptive output feedback controllers for nonlinear compartmental systems with unmodeled dynamics of unknown order. It also guarantees ultimate boundedness of the error signals corresponding to the physical system states as well as the neural network weighting gains.

Nonlinear Adaptive Control for General Anesthesia

Pharmacokinetic Model for Propofol

Almost all anesthetics are *myocardial* depressants, meaning that they decrease the strength of the contraction of the heart and lower cardiac output. As a consequence, decreased cardiac output slows down the transfer of blood from the central compartments comprising the heart, brain, kidney, and liver to the peripheral compartments of muscle and fat. In addition, decreased cardiac output can increase drug concentrations in the central compartments, compounding side effects. This instability can lead to overdosing that, at the very least, can delay recovery from anesthesia and, in the worst case, can result in respiratory and cardiovascular collapse. Alternatively, underdosing can cause psychological trauma from awareness and pain during surgery.

Control of drug effect is clinically important since overdosing or underdosing incur risks for the patient. To illustrate adaptive control for general anesthesia, we consider a hypothetical model for the intravenous anesthetic propofol. The pharmacokinetics of propofol are described by the three-compartment model [49], [52] shown in Figure 4, where x_1 denotes the mass of drug in the central compartment, which is
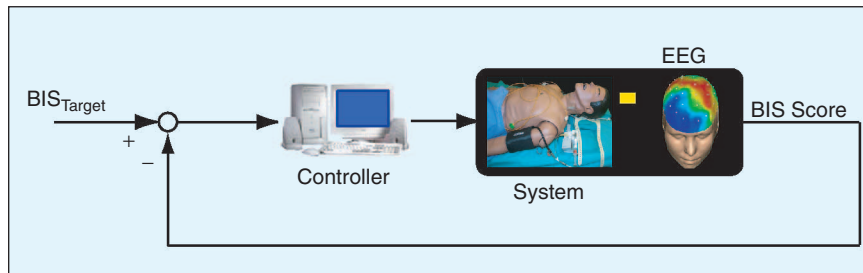

Figure 3. *Adaptive closed-loop control for drug administration. Active control can improve the medical care of patients requiring anesthesia or sedation in the operating room or intensive care unit.*

the site for drug administration and includes the intravascular blood volume as well as *highly perfused* organs (organs with high ratios of blood flow to weight such as the heart, brain, kidney, and liver), which receive a large fraction of the cardiac output. The remainder of the drug in the body is assumed to reside in two peripheral compartments: one identified with muscle and one with fat. The masses in these compartments are denoted by x_2 and x_3, respectively. These compartments receive less than 20% of the cardiac output.

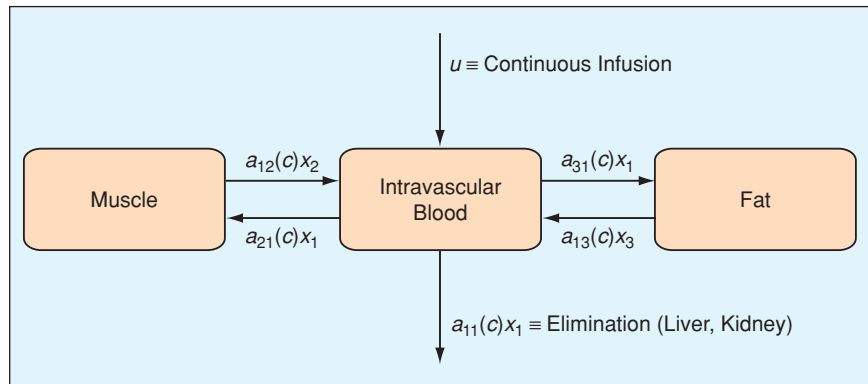A mass balance for the three-state compartmental model is of the form

Figure 4. *Pharmacokinetic model for drug distribution during anesthesia. The central compartment, which is the site for drug administration, is comprised of the intravascular blood volume as well as the highly perfused organs. The two peripheral compartments comprised of muscle and fat receive a small portion of the cardiac output.*

$$\dot{x}_1(t) = -[a_{11}(c(t)) + a_{21}(c(t)) + a_{31}(c(t))]x_1(t)$$
$$+ a_{12}(c(t))x_2(t) + a_{13}(c(t))x_3(t) + u(t),$$
$$x_1(0) = x_{10}, \quad t \geq 0,$$
$$\dot{x}_2(t) = a_{21}(c(t))x_1(t) - a_{12}(c(t))x_2(t), \quad x_2(0) = x_{20},$$
$$\dot{x}_3(t) = a_{31}(c(t))x_1(t) - a_{13}(c(t))x_3(t), \quad x_3(0) = x_{30},$$

where $c(t) = x_1(t)/V_c$, V_c is the volume of the central compartment (about 15 l for a 70-kg patient), $a_{ij}(c)$ for $i \neq j$ is the rate of transfer of drug from the jth compartment to the ith compartment, $a_{11}(c)$ is the rate of drug metabolism and elimination (metabolism typically occurs in the liver), and $u(t)$ is the infusion rate of the anesthetic drug propofol into the central compartment. The transfer coefficients are assumed to be functions of the drug concentration c since it is well known that the pharmacokinetics of propofol are influenced by cardiac output [53]. In turn, cardiac output is influenced by propofol plasma concentrations, both due to *venodilation* (pooling of blood in dilated veins) [54] and myocardial depression [55].

Experimental data indicate that the transfer coefficients a_{ij} are nonincreasing functions of the propofol concentration [54], [55]. The most widely used empirical models for pharmacodynamic concentration-effect relationships are modifications of the Hill equation (4). Applying the Hill equation to the relationship between transfer coefficients and drug concentration implies that

$$a_{ij}(c) = A_{ij}Q_{ij}(c), \quad Q_{ij}(c) = Q_0 C_{50,ij}^{\alpha_{ij}} \big/ \left(C_{50,ij}^{\alpha_{ij}} + c^{\alpha_{ij}} \right),$$

where, for distinct $i, j \in \{1, 2, 3\}$, $C_{50,ij}$ is the drug concentration associated with a 50% decrease in the transfer coef-

ficient, α_{ij} is a parameter that determines the steepness of the concentration-effect relationship, and A_{ij} are positive constants. Note that α_{ij} and A_{ij} are functions of i and j, meaning that there are distinct Hill equations for each transfer coefficient. Furthermore, since for many drugs the rate of metabolism $a_{11}(c)$ is proportional to the rate of drug transport to the liver, we assume that $a_{11}(c)$ is proportional to the cardiac output so that $a_{11}(c) = A_{11}Q_{11}(c)$.

To illustrate the adaptive control of propofol, we assume for simplicity that $C_{50,ij}$ and α_{ij} are independent of i and j. Also, since decreases in cardiac output are observed at clinically utilized propofol concentrations, we arbitrarily assign C_{50} a value of 4 μg/ml, which is in the mid-range of clinically utilized values. We also set $\alpha = 3$, which is typical for ligand-receptor binding (see the discussion in [56]). The nonnegative transfer and loss coefficients A_{12}, A_{21}, A_{13}, A_{31}, and A_{11}, and the parameters $\alpha > 1$, $C_{50} > 0$, and $Q_0 > 0$ are uncertain due to patient gender, weight, preexisting disease, age, and concomitant medication.

Pharmacodynamics and the Effect-Site Compartment

Although propofol concentration in the blood is correlated with lack of responsiveness [57], the concentration cannot be measured in real time during surgery. Since we are more interested in drug effect (depth of hypnosis) than drug concentration, we consider a model involving pharmacokinetics and pharmacodynamics for controlling consciousness. We use an EEG signal, specifically the BIS signal, to access the effect of anesthetic compounds on the brain [40], [58], [59]. Furthermore, we utilize the modified Hill equation (7) to model the relationship between the BIS signal and the effect-site concentration.

The effect-site compartment concentration is related to the concentration in the central compartment by the first-order model
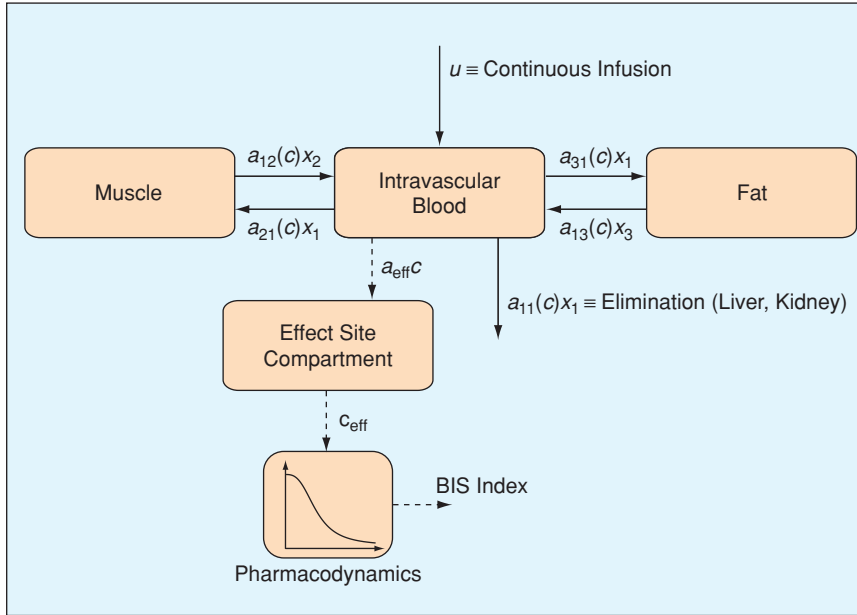
Figure 5. *Pharmacokinetic/pharmacodynamic model with effect-site (brain) compartment. The effect-site compartment is introduced to account for finite equilibration time between the central compartment concentration and the central nervous system concentration.*
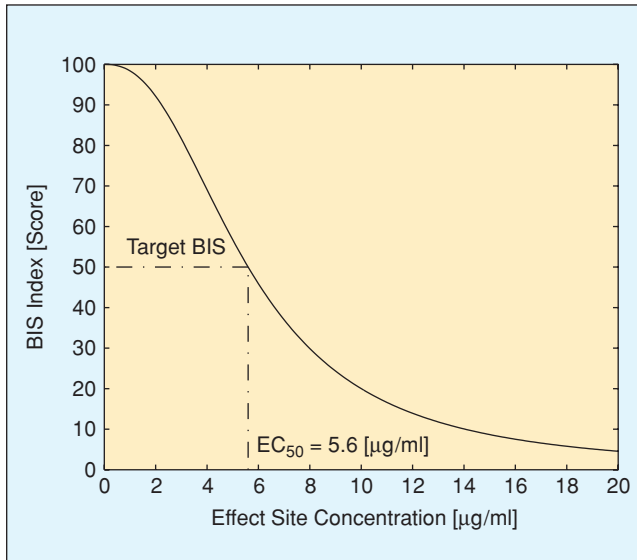

Figure 6. *BIS index (electroencephalogram indicator) versus effect-site concentration. BIS index values of 0 and 100 correspond, respectively, to an isoelectric EEG signal and an EEG signal of a fully conscious patient, while the range 40 to 60 indicates a moderate hypnotic state.*

$$\dot{c}_{eff}(t) = a_{eff}(c(t) - c_{eff}(t)), \quad c_{eff}(0) = c(0), \quad t \geq 0,$$

where a_{eff} in min^{-1} is a time constant. In reality, the effect-site compartment equilibrates with the central compartment in a few minutes. The parameters a_{eff}, EC$_{50}$, and γ are

determined by data fitting and vary from patient to patient. BIS index values of 0 and 100 correspond, respectively, to an *isoelectric* EEG signal (no cerebral electrical activity) and an EEG signal of a fully conscious patient. The range 40 to 60 indicates a moderate hypnotic state [58]. Figure 5 shows the combined pharmacokinetic/pharmacodynamic model for the distribution of propofol.

Simulation

In the following simulation involving the infusion of the anesthetic drug propofol, we set EC$_{50} = 5.6$ μg/ml, $\gamma = 2.39$, and BIS$_0 = 100$, so that the BIS signal is shown in Figure 6. The desired target BIS value BIS$_{target}$ is set at 50. Furthermore, we assume that the effect-site compartment equilibrates instantaneously with the central compartment; that is, we assume that $c_{eff}(t) = c(t)$ for all $t \geq 0$. The adaptive feedback controller for the linearized pharmacodynamic model of (7) derived in [49] is given by

$$u_1(t) = \max\{0, \hat{u}_1(t)\}, \tag{8}$$

where

$$\hat{u}_1(t) = -k_1(t)\left(\text{BIS}(c_{eff}(t)) - \text{BIS}_{target}\right) + \phi_1(t), \tag{9}$$

and $k_1(t)$ and $\phi_1(t)$ are scalars for $t \geq 0$. The update laws $\dot{k}_1(t)$ and $\dot{\phi}_1(t)$ are given by

$$\dot{k}_1(t) = \begin{cases} 0, & \text{if } \hat{u}_1(t) \leq 0, \\ -q_{\text{BIS}_1}(\text{BIS}(c_{eff}(t)) - \text{BIS}_{target})^2, & \text{otherwise,} \end{cases} \tag{10}$$

$$\dot{\phi}_1(t) = \begin{cases} 0, & \text{if } \phi_1(t) = 0 \text{ and } \text{BIS}(c_{eff}(t)) \\ & > \text{BIS}_{target}, \text{ or if } \hat{u}_1(t) \leq 0, \\ \\ \hat{q}_{\text{BIS}_1}(\text{BIS}(c_{eff}(t)) & \text{otherwise,} \\ \quad -\text{BIS}_{target}), \end{cases} \tag{11}$$

where $k_1(0) \leq 0$, $\phi_1(0) \geq 0$, and q_{BIS_1} and $\hat{q}_{\text{BIS}_1}$ are positive constants. Theorem 3.1 of [49] guarantees that

$\text{BIS}(c_{\text{eff}}(t)) \to \text{BIS}_{\text{target}}$ as $t \to \infty$ for all nonnegative values of the pharmacokinetic transfer and loss coefficients $A_{12}, A_{21}, A_{13}, A_{31}, A_{11}$ as well as for all nonnegative coefficients α, C_{50}, and Q_0.

For simulation, we assume $V_c = (0.228 \text{ l/kg})(M \text{ kg})$, where $M = 70$ kg is the mass of the patient, $A_{21}Q_0 = 0.112$ min^{-1}, $A_{12}Q_0 = 0.055$ min^{-1}, $A_{31}Q_0 = 0.0419$ min^{-1}, $A_{13}Q_0 = 0.0033$ min^{-1}, $A_{11}Q_0 = 0.119$ min^{-1}, $\alpha = 3$, and
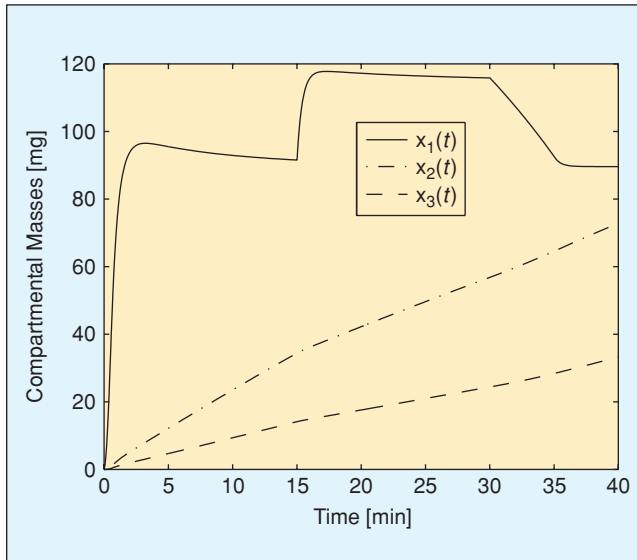
Figure 7. *Compartmental masses versus time. The pharmacodynamic parameters are switched from their nominal values at t = 15 min and back at t = 30 min.*

$C_{50} = 4 \mu\text{g/ml}$ [52]. Furthermore, to illustrate the adaptive controller, we switch the pharmacodynamic parameters EC_{50} and γ, respectively, from 5.6 μg/ml and 2.39 to 7.2 μg/ml and 3.39 at $t = 15$ min and back to 5.6 μg/ml and 2.39 at $t = 30$ min. With $q_{\text{BIS}_1} = 1 \times 10^{-6}$ g/min^2, $\hat{q}_{\text{BIS}_1} = 1 \times 10^{-3}$ g/min^2, and initial conditions $x(0) = [0, 0, 0]^{\text{T}}$ g, $k_1(0) = 0$ g/min, and $\phi_1(0) = 0.01$ g/min, Figure 7 shows the mass of propofol in each of the three compartments versus time. Figure 8 shows the BIS index and the control signal (propofol infusion rate) versus time. Finally, Figure 9 shows the adaptive gain history versus time.

Unlike previous algorithms for closed-loop control of anesthesia [37], [42], the adaptive controller (8)–(11) does not require knowledge of the pharmacokinetic and pharmacodynamic parameters. However, the adaptive controller (8)–(11) does not account for time delays due to equilibration between the central circulation and the effect-site compartment or due to the proprietary signal-averaging algorithm within the BIS monitor. The adaptive controller also ignores measurement noise. Extensive clinical testing is needed to access the significance of these assumptions and approximations. Since there is often a substantial delay between observed changes in patient status and a change in the BIS signal, other measures of depth of anesthesia may be required [60].

Clinical Trials

We have begun clinical studies of the adaptive controller (8)–(11) at the Northeast Georgia Medical Center. In initial clinical testing, we implemented (8)–(11) using a Dell
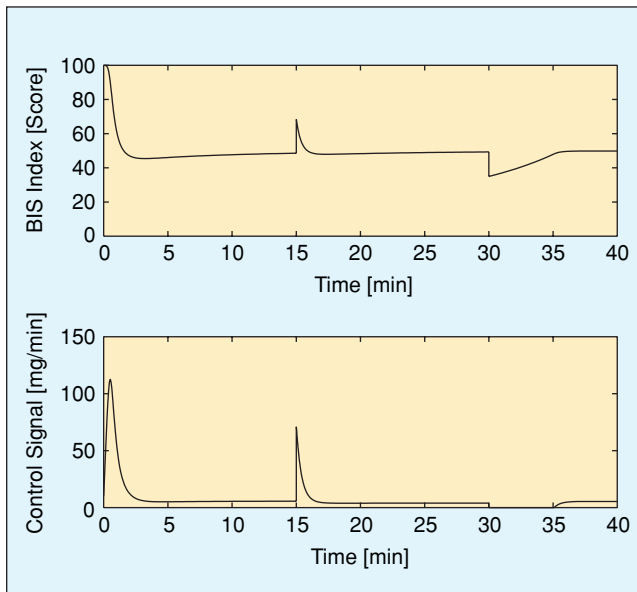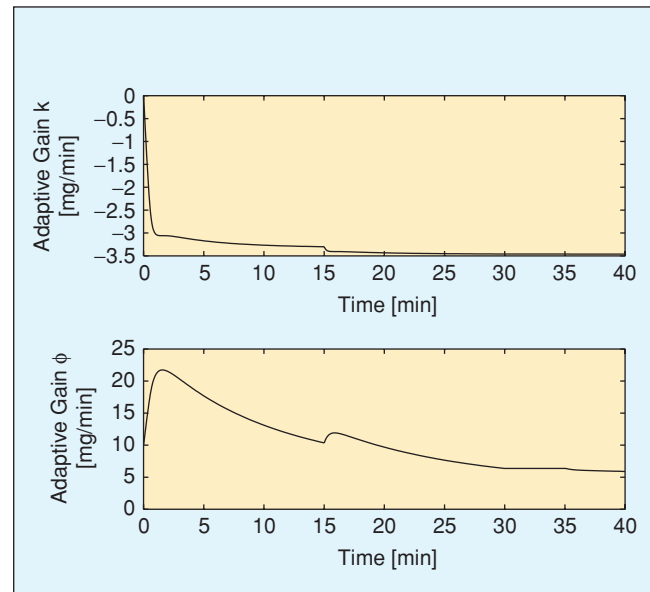

Figure 8. *BIS index versus time and control signal (infusion rate) versus time achieved by the adaptive controller. The adaptive controller does not require knowledge of the pharmacokinetic and pharmacodynamic system parameters.*

Figure 9. *Adaptive gain history versus time. Although the dynamic gains can assume negative values, the physical system states and the control signal are guaranteed to remain nonnegative.*

Latitude C610 laptop computer with a Pentium (R) III processor running under Windows XP, an Aspect A 2000 BIS monitor (rev 3.23), and a Harvard PHD 2000 programmable research pump. The BIS monitor sends a data stream that is updated every 5 s. This data stream contains the BIS signal as well as other parameters such as date, time, signal quality indicator, raw EEG information, and electromyographic data. The data are sent to the serial port of the laptop computer.

The infusion rate $u(t)$ is calculated using a forward Euler method to update the adaptive gains $k_1(t)$ and $\phi_1(t)$ every 0.5 s, using the BIS signal. The infusion rate is communicated to the infusion pump using a 9,600 bpm, eight data bits, two stop bits, and zero parity protocol with the aid of a USB serial port adaptor. An updated infusion rate is sent to the pump at 1-s intervals. Pharmacokinetic simulations predict that a pump update every 5 s or less is adequate in the context of the algorithm under evaluation. An update interval of 1 s was selected in anticipation that future algorithms might benefit from the faster update rate.

The adaptive control algorithm was programmed in Java, an object-oriented programming language chosen for its multiplatform portability tools for rapid prototyping. The program is organized into five modules, which include bisloader, bislogger, controller, pumplogger, and pumploader. Bisloader and bislogger handle communication between the BIS monitor and the computer, while pumploader and pumplogger manage the Harvard pump apparatus. The module bisloader finds the serial port that receives the BIS signal by using the Java class CommPort Identifier; it then invokes bislogger. Bislogger uses the Java class SerialPort EventListener to read the sig-

nal and uses the class StringTokenizer to parse the BIS signal from the input stream. The infusion rate is calculated by the controller module. Finally, pumploader opens the serial port communication to the pump and establishes the communication protocol, while pumplogger delivers the infusion rate to the pump.

The protocol for clinical evaluation of the system was approved by the Institutional Review Board of Northeast Georgia Medical Center. Patients are enrolled after providing informed consent. Our protocol excludes patients requiring emergency surgery, pediatric patients, hemodynamically unstable patients, and patients for whom we anticipate difficult airway management. Otherwise, all elective surgical patients who can provide informed consent are candidates. Preoperative management, including administration of anti-anxiolytic drugs, is left to the discretion of the attending anesthesiologist. Propofol is delivered using the BIS-computer-pump system with a target value of 55. In addition to propofol, all patients receive infusions of either sufentanil or fentanyl with loading doses of 2 μg/ml or 0.25 μg/ml as well as continuous infusions of 2 (μg/ml)/hr or 0.25 (μg/ml)/hr, respectively, to provide analgesia. To ensure patient safety, an independent anesthesia provider observes the progress of the study and can terminate the study if it appears that the patient's safety is being jeopardized by either overdosing or underdosing of propofol.

To date, we have performed 11 clinical trials. For the first clinical trial, Figures 10 and 11 show the controlled BIS index and the control signal versus time. These results are typical of all 11 clinical trials. We have consistently observed an overshoot at the induction of anesthesia with the BIS signal dropping below the target. We suspect that the initial target
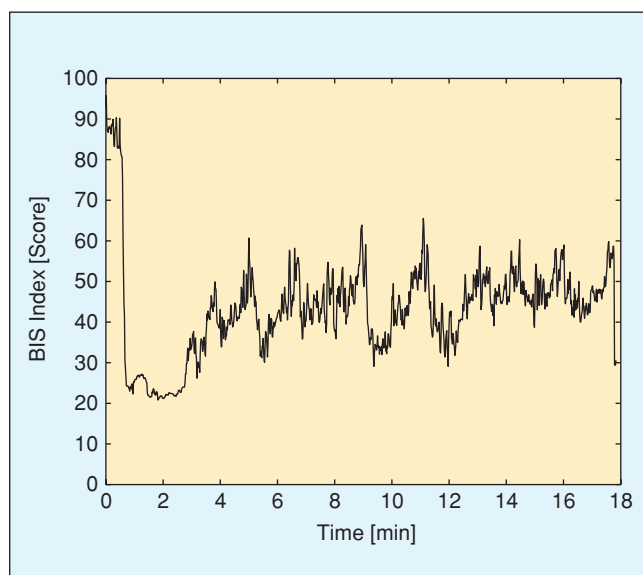

Figure 10. *Controlled BIS index versus time for the first clinical trial. The oscillations are due to measurement noise in the BIS signal.*
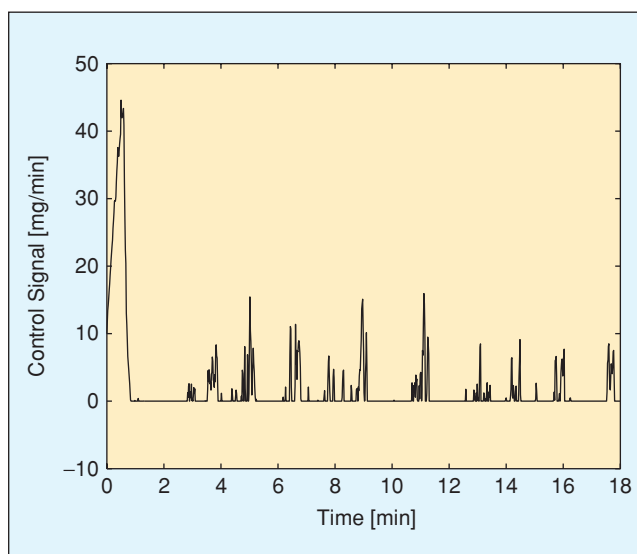

Figure 11. *Infusion rate versus time for the first clinical trial. No initial propofol dose is administered by the anesthesiologist. Induction of anesthesia is delivered by the adaptive controller.*

overshoot stems from the assumption of a linear pharmacodynamic model, which is an approximation. Also, signal averaging time delays within the BIS monitor may contribute to this overshoot. We are preparing clinical trials of an extension of the adaptive controller (8)–(11) that addresses nonlinear pharmacodynamics and time delays [61], [62].

Challenges and Opportunities in Pharmacological Control

Closed-loop control for clinical pharmacology is in its infancy, with numerous challenges and opportunities ahead. An implicit assumption of the control frameworks discussed in this article is that the control law is implemented without regard to actuator amplitude and rate constraints. In pharmacological applications, drug infusion rates vary from patient to patient. To avoid overdosing, it is vital that the infusion rate does not exceed the patient-specific threshold values. As a consequence, actuator constraints, or infusion pump rate constraints, need to be implemented in drug delivery systems [62].

Another important issue for future research is measurement noise. In particular, EEG signals can have as much as 10% variation due to noise. For example, the BIS signal may be corrupted by *electromyographic* noise, or signals emanating from muscle rather than the central nervous system. Although electromyographic noise can be minimized by muscle paralysis, there are other sources of measurement noise, such as electrocautery, that need to be accounted for in the control design processes.

In pharmacokinetic and pharmacodynamic models, the assumption of instantaneous mixing between compartments is not valid. For example, if a bolus of drug is injected, there is a time lag before the drug is detected in the extracellular and intercellular space of an organ [2]. Phase lag due to mixing times can be approximated by including additional compartments in series. To describe the distribution of pharmacological agents in the human body, information on the past system states can be modeled by delay dynamical systems [63]. This extension necessitates the development of adaptive control algorithms for compartmental systems with unknown time delays [62], [64].

Conclusions

Control system technology has a great deal to offer pharmacology, anesthesia, and critical care medicine. Critical care patients, whether undergoing surgery or recovering in intensive care units, require drug administration to regulate physiological variables such as blood pressure, cardiac output, heart rate, and degree of consciousness. The rate of infusion of each administered drug is critical, requiring constant monitoring and frequent adjustments. Open-loop control by clinical personnel can be tedious, imprecise, time consuming, and sometimes of poor quality. Alternatively, closed-loop control can potentially achieve desirable system performance in the face of the highly uncertain and hostile environment of surgery and the intensive care unit. Since robust and adaptive controllers achieve system performance without excessive reliance on system models, robust and adaptive closed-loop control may potentially improve the quality of medical care.

Closed-loop control for clinical pharmacology can significantly advance our understanding of the effects of pharmacological agents and anesthetics, as well as advance the state-of-the-art in drug delivery systems. In addition to delivering sedation to critically ill patients in an acute care environment, potential applications of closed-loop control include glucose, heart rate, and blood pressure regulation. Payoffs will arise from improvements in medical care, health care, reliability of drug dosing equipment, and reduced health-care costs.

Acknowledgment

References

[1] P.G. Welling, *Pharmacokinetics: Processes, Mathematics, and Applications*, 2nd ed. Washington, DC: American Chemical Soc., 1997.

[2] J.A. Jacquez, *Compartmental Analysis in Biology and Medicine*. Ann Arbor, MI: Univ. of Michigan Press, 1985.

[3] M. Morari and A. Gentilini, "Challenges and opportunities in process control: Biomedical processes," *AIChE J.*, vol. 47, no. 10, pp. 2140–2143, 2001.

[4] A.G. Gilman, J.G. Hardman, L.E. Limbird, *Goodman and Gilman's The Pharmacological Basis of Therapeutics*, 10th ed. New York: McGraw-Hill, 1996.

[5] T.K. Henthorn, T.C. Krejcie, C.U. Niemann, C. Enders-Klein, C.A. Shanks, and M.J. Avram, "Ketamine distribution described by a recirculatory pharmacokinetic model is not stereo-selective," *Anesthesiology*, vol. 91, no. 6, pp. 1733–1743, 1999.

[6] H. Schwilden, "A general method for calculating the dosage scheme in linear pharmacokinetics," *Eur. J. Clin. Pharmacol.*, vol. 20, no. 5, pp. 379–386, 1981.

[7] M. Davidian and D.M. Giltinan, *Nonlinear Models for Repeated Measurement Data*. Boca Raton, FL: Chapman and Hall, 1995.

[8] L.B. Sheiner and S.L. Beal, "Evaluation of methods for estimating population pharmacokinetic parameters II. Biexponential model and experimental pharmacokinetic data," *J. Pharmacokinetics Biopharm.*, vol. 9, no. 5, pp. 635–651, 1981.

[9] L.B. Sheiner, "The population approach to pharmacokinetic data analysis: Rationale and standard data analysis methods," *Drug Metabolism Rev.*, vol. 15, no. 1–2, pp. 153–171, 1984.

[10] A.V. Hill, "The possible effects of the aggregation of the molecules of haemoglobin on its dissociation curves," *J. Phys.*, vol. 40, no. 1, pp. 4–7, 1910.

[11] W. Lu and J.M. Bailey, "The reliability of pharmacodynamic analysis by logistic regression: A computer simulation study," *Anesthesiology*, vol. 92, no. 4, pp. 985–992, 2000.

[12] W. Lu, J.G. Ramsay, and J.M. Bailey, "Reliability of pharmacodynamic analysis by logistic regression: Mixed-effects modeling," *Anesthesiology*, to appear.

[13] J.M. Alvis, J.G. Reves, A.V. Govier, P.G. Menkhaus, C.E. Henling, J.A. Spain, and E. Bradley, "Computer-assisted continuous infusions of fentanyl during cardiac anesthesia: Comparison with a manual method," *Anesthesiology*, vol. 63, no. 1, pp. 41–49, 1985.

[14] S.L. Shafer, J.R. Varvel, N. Aziz, and J.C. Scott, "Pharmacokinetics of fentanyl administered by computer-controlled infusion pump," *Anesthesiology*, vol. 73, no. 6, pp. 1092–1102, 1990.

[15] J.M. Bailey and S.L. Shafer, "A simple analytical solution to the three compartment pharmacokinetic model suitable for computer controlled infusion pumps," *IEEE Trans. Biomed. Eng.*, vol. 38, no. 6, pp. 522–525, 1991.

[16] S.L. Shafer and K. Gregg, "Algorithms to rapidly achieve and maintain stable drug concentrations at the site of drug effect with a computer-controlled infusion," *J. Pharmacokinetics Biopharm.*, vol. 20, no. 2, pp. 147–169, 1992.

[17] P.O. Maitre and D.R. Stanski, "Bayesian forecasting improves the prediction of intraoperative plasma concentrations of alfentanil," *Anesthesiology*, vol. 69, no. 5, pp. 652–659, 1988.

[18] J.H. Levy, L.G. Michelsen, J.S. Shanewise, J.M. Bailey, and J.G. Ramsay, "Postoperative cardiovascular management," in *Cardiac Anesthesia*, 4th ed., J. Kaplan, Ed. Philadelphia, PA: Saunders, 1999.

[19] J. Slate, "Model-based design of a controller for infusion of sodium nitroprusside during postsurgical hypertension," Ph.D. dissertation. Univ. of Wisconsin-Madison, Madison, WI, 1980.

[20] L.C. Sheppard, "Computer control of the infusion of vasoactive drugs," *Ann. Biomed. Eng.*, vol. 8, pp. 431–444, 1980.

[21] J.M. Arnsparger, B.C. McInnis, J.R. Glover, and N.A. Normann, "Adaptive control of blood pressure," *IEEE Trans. Biomed. Eng.*, vol. 30, pp. 168–176, 1983.

[22] W.G. He, H. Kaufman, and R.J. Roy, "Multiple model adaptive control procedure for blood pressure control," *IEEE Trans. Biomed. Eng.*, vol. 33, no. 1, pp. 10–19, 1986.

[23] R.R. Rao, C.C. Palerm, B. Aufderheide, and B.W. Bequette, "Automated regulation of hemodynamic variables," *IEEE Eng. Med. Biol. Mag.*, vol. 20, no. 1, pp. 24–38, 2001.

[24] R.R. Rao, B. Aufderheide, and B.W. Bequette, "Experimental studies on multiple-model predictive control for automated regulation of hemodynamic variables," *IEEE Trans. Biomed. Eng.*, vol. 50, no. 3, pp. 277–288, 2003.

[25] G.A. Pajunen, M. Steinmetz, and R. Shankar, "Model reference adaptive control with constraints for postoperative blood pressure management," *IEEE Trans. Biomed. Eng.*, vol. 37, no. 7, pp. 679–687, 1990.

[26] C.T. Chen, W.L. Lin, and T.S. Kuo, "Adaptive control of arterial blood pressure with a learning controller based on multilayer neural networks," *IEEE Trans. Biomed. Eng.*, vol. 44, no. 7, pp. 601–609, 1997.

[27] K. Behbehani and R.R. Cross, "A controller for regulation of mean arterial blood pressure using optimum nitroprusside infusion rate," *IEEE Trans. Biomed. Eng.*, vol. 38, no. 6, pp. 513–521, 1991.

[28] C. Yu, R.J. Roy, H. Kaufman, and B.W. Bequette, "Multiple-model adaptive predictive control of mean arterial pressure and cardiac output," *IEEE Trans. Biomed. Eng.*, vol. 39, no. 8, pp. 765–778, 1992.

[29] C.M. Held and R.J. Roy, "Multiple drug hemodynamic control by means of a supervisory-fuzzy rule-based adaptive control system: Validation on a model," *IEEE Trans. Biomed. Eng.*, vol. 42, no. 4, pp. 371–385, 1995.

[30] E.I. Eger, *Anesthetic Uptake and Action*. Baltimore, MD: Williams & Wilkins, 1997.

[31] I.J. Rampil, "A primer for EEG signal processing in anesthesia," *Anesthesiology*, vol. 89, no. 4, pp. 980–1002, 1998.

[32] J.A. Ross, R.T. Wloch, D.C. White, and D.W. Hawes, "Servo-controlled closed-circuit anaesthesia. A method for the automatic control anaesthesia produced by a volatile agent in oxygen," *Br. J. Anaesth.*, vol. 55, no. 11, pp. 229–230, 1982.

[33] R.G. Ritchie, E.A. Ernst, B.L. Pate, J.D. Pearson, and L.C. Shepherd, "Closed-loop control of an anesthesia delivery system: Development and animal testing," *IEEE Trans. Biomed. Eng.*, vol. 34, no. 6, pp. 437–443, 1987.

[34] R. Vishnoi and R.J. Roy, "Adaptive control of closed-circuit anesthesia," *IEEE Trans. Biomed. Eng.*, vol. 38, no. 1, pp. 39–47, 1991.

[35] G.I. Jee and R.J. Roy, "Adaptive control of multiplexed closed-circuit anesthesia," *IEEE Trans. Biomed. Eng.*, vol. 39, no. 10, pp. 1071–1080, 1992.

[36] R.G. Bickford, "Automatic electroencephalographic control of anesthesia (servo-anesthesia)," *Electroenceph. Clin. Neurophysiol.*, vol. 3, pp. 83–86, 1951.

[37] H. Schwilden, J. Schuttler, and H. Stoeckel, "Closed-loop feedback control of methohexital anesthesia by quantitative EEG analysis in humans," *Anesthesiology*, vol. 67, no. 3, pp. 341–347, 1987.

[38] H. Schwilden, H. Stoeckel, and J. Schuttler, "Closed-loop feedback control of propofol anesthesia by quantitative EEG analysis in humans," *Br. J. Anaesth.*, vol. 62, no. 3, pp. 290–296, 1989.

[39] P.S. Sebel, E. Lang, I.J. Rampil, P. White, R.C.M. Jopling, N.T. Smith, P.S. Glass, and P. Manberg, "A multicenter study of bispectral electroencephalogram analysis for monitoring anesthetic effect," *Anesth. Analg.*, vol. 84, no. 4, pp. 891–899, 1997.

[40] P.S. Glass, M. Bloom, L. Kearse, C. Rosow, P. Sebel, and P. Manberg, "Bispectral analysis measures sedation and memory effects of propofol, midazolam, isoflurane, and alfentanil in normal volunteers," *Anesthesiology*, vol. 86, no. 4, pp. 836–847, 1997.

[41] T.W. Schnider, C.F. Minto, and D.R. Stanski, "The effect compartment concept in pharmacodynamic modelling," *Anaes. Pharmacol. Rev.*, vol. 2, pp. 204–213, 1994.

[42] M. Struys, T. De Smet, L. Versichelen, S. Vd Vilde, R. Vd Broecke, and E. Mortier, "Comparison of closed-loop controlled administration of propofol using BIS as the controlled variable versus 'standard practice' controlled administration," *Anesthesiology*, vol. 95, no. 1, pp. 6–17, 2001.

[43] P.S.A. Glass and I.J. Rampil, "Automated anesthesia: Fact or fantasy?," *Anesthesiology*, vol. 95, no. 1, pp. 1–2, 2001.

[44] A.R. Absalom, N. Sutcliffe, and G.N. Kenny, "Closed-loop control of anesthesia using bispectral index: Performance assessment in patients undergoing major orthopedic surgery under combined general and regional anesthesia," *Anesthesiology*, vol. 96, no. 1, pp. 67–73, 2002.

[45] G.N. Kenny and H. Mantzardis, "Closed-loop control of propofol anaesthesia," *Br. J. Anaesth.*, vol. 83, no. 2, pp. 223–228, 1999.

[46] D.A. Linkens, M.F. Abbod, and J.E. Peacock, "Clinical implementation of advanced control in anaesthesia," *Trans. Inst. Meas. Contr.*, vol. 22, no. 4, pp. 303–330, 2000.

[47] A. Gentilini, M. Rossoni-Gerosa, C.W. Frei, R. Wymann, M. Morari, A.M. Zbinden, and T.W. Schnider, "Modeling and closed-loop control of hypnosis by means of bispectral index (BIS) with isoflurane," *IEEE Trans. Biomed. Eng.*, vol. 48, no. 8, pp. 874–889, 2001.

[48] W.M. Haddad, T. Hayakawa, and J.M. Bailey, "Adaptive control for nonnegative and compartmental dynamical systems with applications to general anesthesia," *Int. J. Adapt. Contr. Signal Processing*, vol. 17, no. 3, pp. 209–235, 2003.

[49] W.M. Haddad, T. Hayakawa, and J.M. Bailey, "Nonlinear adaptive control for intensive care unit sedation and operation room hypnosis," in *Proc. Amer. Control Conf.*, Denver, CO, June 2003, pp. 1808–1813.

[50] T. Hayakawa, W.M. Haddad, J.M. Bailey, and N. Hovakimyan, "Passivity-based neural network adaptive output feedback control for nonlinear nonnegative dynamical systems," in *Proc. IEEE Conf. Decision and Control*, Maui, HI, Dec. 2003, pp. 5697–5702.

[51] T. Hayakawa, W.M. Haddad, N. Hovakimyan, and J.M. Bailey, "Neural network adaptive dynamic output feedback control for nonlinear nonnegative systems using tapped delay memory units," in *Proc. Amer. Control Conf.*, Boston, MA, July 2004, pp. 4505–4510.

[52] B. Marsh, M. White, N. Morton, and G.N. Kenny, "Pharmacokinetic model driven infusion of propofol in children," *Br. J. Anaesth.*, vol. 67, no. 1, pp. 41–48, 1991.

[53] R.N. Upton, G.I. Ludrook, C. Grant, and A. Martinez, "Cardiac output is a determinant of the initial concentration of propofol after short-term administration," *Aneth. Analg.*, vol. 89, no. 3, pp. 545–552, 1999.

[54] M. Muzi, R.A. Berens, J.P. Kampine, and T.J. Ebert, "Venodilation contributes to propofol-mediated hypotension in humans," *Aneth. Analg.*, vol. 74, no. 6, pp. 877–883, 1992.

[55] E.F. Ismail, S.J. Kim, and M.R. Salem, "Direct effects of propofol on myocardial contractility in situ canine hearts," *Anesthesiology*, vol. 79, no. 5, pp. 964–972, 1992.

[56] R.G. Eckenhoff and J.S. Johansson, "On the relevance of 'clinically relevant concentrations' of inhaled anesthetics in in vitro experiments," *Anesthesiology*, vol. 91, no. 3, pp. 856–860, 1999.

[57] T. Kazama, K. Ikeda, and K. Morita, "The pharmacodynamic interaction between propofol and fentanyl with respect to the suppression of somatic or hemodynamic responses to skin incision, peritoneum incision, and abdominal wall retraction," *Anesthesiology*, vol. 89, no. 4, pp. 894–906, 1998.

[58] J.C. Sigl and N.G. Chamoun, "An introduction to bispectral analysis for the electroencephalogram," *J. Clin. Monit.*, vol. 10, no. 6, pp. 392–404, 1994.

[59] E. Mortier, M. Struys, T. De Smet, L. Versichelen, and G. Rolly, "Closed-loop controlled administration of propofol using bispectral analysis," *Anaesthesia*, vol. 53, no. 8, pp. 749–754, 1998.

[60] X.S. Zhang, R.J. Roy, and E.W. Jensen, "EEG Complexity as a measure of depth of anesthesia for patients," *IEEE Trans. Biomed. Eng.*, vol. 48, no. 12, pp. 1424–1433, 2001.

[61] W.M. Haddad, J.M. Bailey, T. Hayakawa, and N. Hovakimyan, "Neural network adaptive output feedback control for intensive care unit sedation and intraoperative anesthesia," submitted for publication.

[62] Q. Hui, W.M. Haddad, V. Chellaboina, and T. Hayakawa, "Adaptive control of mammillary drug delivery systems with actuator amplitude constraints and system time delays," submitted for publication.

[63] W.M. Haddad and V. Chellaboina, "Stability and dissipativity theory for nonnegative and compartmental dynamical systems with time delay," in *Advances in Time-Delay Systems*, S.I. Niculescu and K. Gu, Eds. New York: Springer, 2004, pp. 421–435.

[64] V. Chellaboina, W.M. Haddad, J. Ramakrishnan, and T. Hayakawa, "Direct adaptive control of nonnegative and compartmental systems with time delay," in *Proc. Amer. Control Conf.*, Boston, MA, July 2004, pp. 1235–1240.

James M. Bailey received the B.S. degree from Davidson College in 1969, the Ph.D. in chemistry (physical) from the University of North Carolina at Chapel Hill in 1973, and the M.D. degree from Southern Illinois University School of Medicine in 1982. He was a Helen Hay Whitney Fellow at the California Institute of Technology from 1973–1975 and assistant professor of chemistry and biochemistry at Southern Illinois University from 1975–1979. After receiving his M.D. degree, he completed a residency in anesthesiology and then a fellowship in cardiac anesthesiology at the Emory University School of Medicine affiliated hospitals. From 1986–2002, he was an assistant professor of anesthesiology and then associate professor of anesthesiology at Emory, where he also served as director of the critical care service. In September 2002, he moved his clinical practice to Northeast Georgia Medical Center in Gainesville, Georgia, as director of cardiac anesthesia and consultant in critical care medicine. He remains affiliated with the Emory University School of Medicine Department of Anesthesiology as a clinical associate professor. He is board certified in anesthesiology, critical care medicine, and transesophageal echocardiography. His research interests are focused on pharmacokinetic and pharmacodynamic modeling of anesthetic and vasoactive drugs and, more recently, applications of control theory in medicine. He is the author or coauthor of 98 journal articles, conference publications, and book chapters.

Wassim M. Haddad (wm.haddad@aerospace.gatech.edu) received the B.S., M.S., and Ph.D. degrees in mechanical engineering from the Florida Institute of Technology, Melbourne, in 1983, 1984, and 1987, respectively, with specialization in dynamical systems and control. From 1987 to 1994, he was a consultant for the Structural Controls Group of the Government Aerospace Systems Division, Harris Corporation, Melbourne, Florida. In 1988, he joined the faculty of the Mechanical and Aerospace Engineering Department at the Florida Institute of Technology, where he founded and developed the systems and control option within the graduate program. Since 1994, he has been a member of the faculty in the School of Aerospace Engineering at the Georgia Institute of Technology, where he is a professor. His research contributions in linear and nonlinear dynamical systems and control are documented in over 400 archival journal and conference publications. His recent research is concentrated on nonlinear robust and adaptive control, nonlinear dynamical system theory, large-scale systems, hierarchical nonlinear switching control, hybrid and impulsive control for nonlinear systems, system thermodynamics, thermodynamic modeling of mechanical and aerospace systems, nonlinear analysis and control for biological and physiological systems, and active control for clinical pharmacology. He is an NSF Presidential Faculty Fellow, a member of the Academy of Nonlinear Sciences, and a coauthor of *Hierarchical Nonlinear Switching Control Design with Applications to Propulsion Systems* (Springer-Verlag, 2000) and *Thermodynamics: A Dynamical Systems Approach* (Princeton University Press, 2005). He can be contacted at the School of Aerospace Engineering, Georgia Institute of Technology, Atlanta, GA 30332-0150 USA.