Applying Pharmacokinetic and Pharmacodynamic Models in the Operating Room: Validation of Response Surface Models

F. H. Sakaguchi

***Abstract*—Pharmacokinetics are used to model drug concentrations in the body. These predictions can be combined with pharmacodynamic response surface models that predict the effect of multiple drugs acting on the body. This study combined several pharmacokinetic and pharmacodynamic models to predict "adequate anesthesia." These predictions were compared to observations in patients. While these specific model combinations are not accurate predictors of anesthesia for the recovery of responsiveness and tracheal intubation, a few combinations are reasonable predictors of the loss of responsiveness and also for the analgesia necessary for the first skin incision. The Schnider propofol model and a fentanyl scaling factor of 1.2 are empirically the most accurate PK models in combination with the pharmacodynamic models used.**

I. Introduction

Response surfaces have been used to describe the pharmacological effect of two drugs at once. The x and y-axes are typically the independent variables in drug concentration units. The z-axis is typically the predicted effect of the drugs in combination. In the realm of anesthesia, the anesthetic concentrations can be estimated by pharmacokinetic (PK) models. Pharmacodynamic (PD) models predict the level of anesthetic effect as a function of drug concentration. This observational study evaluates how accurately a combination of pharmacokinetic and pharmacodynamic models predicted the level of anesthesia in 24 patients undergoing abdominal laproscopic surgery.

II. Methods

A. Study Design

This observational study was designed to assess the accuracy of combined pharmacokinetic and pharmacodynamic (PK-PD) models in the operating room to predict adequate anesthesia for four clinical milestones: 1) loss of responsiveness, 2) laryngoscopy and tracheal intubation, 3) the first skin incision, and 4) the recovery of responsiveness at the end of surgery. This study was only observational, and had minimal impact on the "standard practice" of the anesthesiologist. Two types of data were collected intraoperatively: 1) drug dosing information and 2) observations of the patient responses and nonresponses at clinical milestones. Comparison of the PK-PD combined model predictions with the patient observations was performed post hoc.

B. Subjects and Observations at Clinical Milestones

With institutional review board approval from the University Hospital and informed consent of the patients, we studied 24, ASA physical status I, II, and III, patients (11 males and 13 females) scheduled for abdominal laparoscopic surgery under total intravenous anesthesia. All patients denied having cardiovascular, hepatic, or renal disease or a history of alcohol or drug abuse. The anesthetic regimen was limited to propofol for sedation, and remifentanil and fentanyl for analgesia.

Propofol and remifentanil syringes were loaded into separate infusion pumps (Medfusion 3010a, Medex, Inc., Dublin, OH, USA). After the patient entered the OR, the primed remifentanil and propofol infusion lines were attached at the patient's wrist. This decreased the delay of drug delivery by minimizing the tubing deadspace flushed by the IV drip. The anesthetists administered drug boluses for both induction and maintenance through the second IV access port distal from the patient. An intra-lab software interface collected data from the two infusion pumps. A research nurse and a graduate student observer recorded drug boluses given manually.

C. Pharmacokinetic Modeling

The modeling and assessment of the accuracy of the combined PK-PD model predictions occurred post hoc. The PK simulations were run using the patient and drug dosing information. Each drug (propofol, remifentanil, and fentanyl) had its own PK model. The pharmacokinetics of each drug was assumed independent of the concentration of the other drugs. We selected the Minto-Schnider model for remifentanil [1] and an adapted Shafer *et al.* model for fentanyl [2]. (The keo for the fentanyl model was calculated using the Tpeak [3] from Scott and Stanski's data [4], instead of simply using the keo from their data.) However, selecting the propofol model was not as straight forward.

F. H. Sakaguchi is with the Department of Anesthesiology, University of Utah, Salt Lake City, UT 84132 USA. (phone: 801-879-9461; fax: 240-208-0362; e-mail: farrants@ abl.med.utah.edu).

We wished to use the Tackley model for propofol [5], since it had been used in the target-controlled-infusion system used for building the PD models of Kern *et al.* [6]. However, the Tackley model lacks an effect-site compartment while both the Gepts [7] and the Schnider [8] models include effect-site compartments. The Gepts model lacks the PD data necessary to calculate the keo based on the time to peak effect. Therefore, we adapted the Tackley model to include an effect-site compartment by using the k1e and ke0 variables of the Gepts and Schnider models. Ultimately, four sets of rate constants were used for the propofol model—the Gepts model (G), the Schnider model (S), and the Tackley model using the Gepts effect-site (TG), and the Tackley model using the Schinder effect-site (TS).

D. Pharmacodynamic Modeling

The PD models used the modeled effect-site concentrations (Ceff) at the milestones of 1) LOR, 2) TI, 3) SI, and 4) ROR to predict the level of anesthesia. The PD models predict the percentage of the population at those given drug concentrations that would be "adequately anesthetized" for each given surrogate measure. Because the PK and PD models are independent of one another, they could be "mixed and matched," producing different overall predictions for the depth of anesthesia. The PK-PD combined model predictions were then compared to the actual patient responses and their accuracies were assessed.

The response surfaces explicitly describe the PD interaction between propofol and remifentanil only. To account for the analgesic effect of fentanyl (commonly used in standard practice), we used its relative opioid effect. In other words, the Ceff of fentanyl was scaled by its relative potency to the Ceff of remifentanil. However, the exact relative potency is difficult to ascertain from the literature. Therefore, we chose to consider three different scaling factors: 1 ng/ml of fentanyl exerted a PD effect equal to the effect of 1.2, 1.7, or 2.4 ng/ml of remifentanil [9]. To calculate the total opioid concentration, normalized to remifentanil, the predicted concentrations of fentanyl were multiplied by their scaling factors and were added to the predicted concentration of remifentanil. We estimated the PD effect (the likelihood of adequate anesthesia) from the total opioid Ceff and the propofol Ceff; the total opioid Ceff and propofol Ceff marked a data point on the response surface.

Kern *et al.* used four surrogate measures to predict anesthetic effects of sedation and analgesia [6]; they used 1) the Observer's Assessment of Alertness/Sedation (OAA/S) to predict the depth of hypnosis [10], both 2) responses to electrical tetany and 3) responses to shin algometry to predict the analgesia necessary for skin incision, and 4) responses to laryngoscopy to predict the analgesia necessary for laryngoscopy and endotracheal intubation. Although the OAA/S follows a discrete scale from 1 to 5, Kern *et al.* treated OAA/S scores above and below 4 as binary states of sedation. Using the raw data of Kern *et al.*, we recreated his sedation model using different criteria. We calculated additional response surfaces for sedation with the criteria of OAA/S scores below 3 and 2 using the least-squares method. Kern *et al.* assumed the steady-state plasma concentrations equaled the anesthetic Ceffs. Thus, the Ceff estimates from the PK models were used as inputs into the PD models of Kern *et al.*.

Merten *et al.* observed actual surgical measures for responses to pain [11]. Response surfaces were created for the measures of 1) Laryngoscopy, 2) Endotracheal Intubation, and 3) Awakening. Laryngoscopy was repeated at three different concentration pairs. Both laryngoscopy and endotracheal intubation were performed at pseudo-steady-state concentrations to approximate Ceffs. Similarly, the Recovery of Consciousness was at the end of surgery when the anesthetic was eliminated by the body but was not marked by rapid changes in blood concentrations. Ceff estimates from the PK models were used as inputs for the PD models of Merten *et al.*.

E. Data Analysis

PK simulations used the dosing information gathered in the OR. From the OR patient observations, we gathered the time points of the surgical stimuli. We used the estimated Ceff pairs of propofol and total opioid (normalized to remifentanil) at these time points to calculate PD predictions appropriate for the surgical measure. Each combined PK-PD model was treated as an entirely different scenario, so that the accuracy of each unique combination of PK-PD models could be evaluated for its accuracy relative to the population and the surgical measure. Each prediction was also paired with an actual observed patient state of "responsive" or "nonresponsive."

For a given population PD model, a 50% or 95% prediction of adequate anesthesia defines the anesthetic concentrations needed to attain 100% anesthetic effect in 50% or 95% of the population for that stimulus. These PD models come from pooled individual data. However, when the population model is used to infer information about the individual, the meaning of the prediction changes somewhat. For a PD prediction for an individual, we see what percentage of the population is expected to be anesthetized, and infer the likelihood of anesthesia for an individual. Thus, for sedation PD models, a 50% value describes the concentrations at which 50% of the population is 100% sedated and is taken as the likelihood of an "average" individual being completely sedated. For PD models of noxious stimuli, the PD model predicted level of anesthesia refers to the percentage of the population that would be anesthetized and estimates the likelihood of the patient to not exhibit a heart rate or other somatic response at those anesthetic concentrations.

The analyses are grouped by the clinical measure for which the adequacy of anesthesia was being predicted. For each measure, the different PK and models were combined. All the possible PK-PD model combinations are summarized in Table 1 by the surgical measure of interest.

The LOR and ROR PK-PD combined model predictions could not be tested simply at the LOR and ROR observation times as the recorded data did not include any "false positives" or "false negatives." Therefore, the mean elapsed times of the

TABLE 1
PK-PD MODEL COMBINATIONS TESTED

SURGICAL MEASURE	PK MODEL VARIATIONS: PROPOFOL PK MODELS	PK MODEL VARIATIONS: FENTANYL: REMIFENTANIL RELATIVE POTENCY	PD MODELS	# OF POSSIBLE PK-PD MODEL COMBINATIONS
Loss of Responsiveness			Kern OAA/S 4-3 Kern OAA/S 3-2 Kern OAA/S 2-1 Mertens Awakening	4*3*4 = 48
Laryngoscopy and Endotracheal Intubation	Tackley – Gepts Tackley – Schnider Gepts Schnider	1.2:1 1.7:1 2.4:1	Kern Laryngoscopy Mertens Laryngoscopy Mertens Intubation	4*3*3 = 36
First Skin Incision			Kern Algometry Kern Tetany	4*3*2 = 24
Recovery of Responsiveness			Kern OAA/S 4-3 Kern OAA/S 3-2 Kern OAA/S 2-1 Mertens Awakening	4*3*4 = 48

LOR and ROR observations were used. These elapsed times were calculated from the first drug administered in the operating room until the observed LOR and from the last trocar removal and the beginning of wound closure until the observed ROR. In contrast, the TI and SI data were easily analyzed directly from the recorded observations.

For all surgical measures, we compared the PK-PD predictions to the observations of adequate anesthesia as a population using two different nonparametric tests (the two-tailed Fisher exact test and the P_K test) and by the coefficient of determination (r^2). The two-tailed Fisher exact test was used to evaluate whether the associations between the observations and the PK-PD combined model predictions relative to the 50% and 95% levels of effect or isobols were statistically significant [12]. The P_K is a measure of the accuracy with which the combined PK-PD models predict the observed patient states [13], [14]. r^2 is a common "goodness-of-fit" measure [12].

The two-tailed Fisher exact test analyzes 2x2 contingency tables for small sample sizes by computing the exact probability of that specific arrangement of data from all possible arrangements, with the hypothesis that the rows and columns are independent [12]. The number of "nonanesthetized" and "anesthetized" patients below and above a given isobol were tabulated. To test the sedation LOR and ROR predictions, the two-tailed Fisher exact tests were used at the mean times relative to the 50% and 95% isobols. For TI and SI the Fisher exact tests were used on the PK-PD combined model predictions at the observed stimuli, relative to the 50% and 95% isobols. The p-value for the two-tailed Fisher exact test was calculated using GraphPad InStat (v. 3.05, 32 bit for Win 95/NT, GraphPad Software, San Diego, California, USA, www.graphpad.com). A p-value < 0.005 or 0.05 was considered extremely statistically significant or statistically significant, respectively, and represents the probability of the predictions and the observations as tabulated in the contingency table occurring by chance.

We also used the prediction probability P_K test (a nonparametric rank-order test) to assess the overall accuracy of the model predictions [13], [14]. This test is set up by pairing the PK-PD combined model predictions with the patient state. In total, the P_K statistic was used on four sets of data: 1) the LOR predictions at the mean time point, 2) the ROR predictions at the mean time, 3) the TI data, and 4) the SI data. The P_K was calculated using a macro written for Excel by and available from Warren D. Smith (copyright 1996). When $P_K = 1$, the order of the model predictions always correctly predict the order of the patient state, whereas when $P_K = 0.5$, the models do not predict the patient state.

The coefficient of determination or r^2 (the square of the Pearson product-moment correlation coefficient) is a measure of how much of the variability in the data is predicted by the PK-PD combined models [12]. Again, the mean and median data for the sedation predictions were not lumped. The correlation coefficient was calculated in Excel. An $r^2 = 1$ suggests a perfect correlation while an $r^2 = 0$ suggests no correlation.

III. RESULTS

A. Recruitment and Data Collection

All 24 patients enrolled, completed the study. Data are presented as the mean ± standard deviation. The mean age was 38.9 ± 12.4 years and the mean weight was 86.4 ± 22.6 kg, and the mean height was 171.58 ± 8.99 cm. Of the 24 observed surgical cases, 15 were laproscopic cholecystectomies, 6 surgeries were laproscopic hernia repairs, and 3 were laproscopic nissen fundoplications. The anesthetic was delivered by 2 experienced CRNAs in 16 cases, 2 third year residents in 2 cases, 2 second year residents in 3 cases, and a first year resident in 3 cases.

Although propofol was the only intraoperative sedative, with two exceptions, the patients received midazolam (average dose of 1.61 mg, ± 0.49) prior to entering the operating room (OR). (One of the patients declined midazolam and another

patient received midazolam after arriving in the OR.) Remifentanil and fentanyl were the primary analgesics. Ketorolac tromethamine was also given in 13 cases late in the procedure for maintenance and post-operative pain management. However, neither the kinetics nor the dynamics of ketorolac tromethamine were modeled. The observations of patients at clinical milestones are summarized in Table 2. LOR was indicated during induction when the patient no longer responded to the anesthesiologist. A heart rate response was indicated if there was a 20% increase in the patient heart rate within 1 minute after laryngoscopy and tracheal intubation or the first skin incision. ROR was indicated when the patient obeyed simple commands by the anesthesiologist (such as "Open your eyes," or "Please take a deep breath so we can take out your breathing tube.")

B. Combined PK-PD Model Predictions

The likelihood of adequate anesthesia was calculated using the predicted propofol and total opioid (normalized to remifentanil) Ceff pairs and the appropriate PD models. For a single patient at each clinical milestone, there were many different PK-PD combined model predictions, summarized in Table 2. The best performing PK-PD combined model estimates are described below under Statistical Analysis. The results of all 156 PK-PD model combinations for all clinical measures are available from the author upon request.

C. Statistical Analysis

For LOR and using the two-tailed Fisher exact test, all 48 of the PK-PD combinations resulted in statistically significant associations between the observations and model predictions relative to both 50% and 95% isobols at the mean LOR time. These results, with the specific models, are shown in Table 3, along with the top three performing model combinations at the mean and median times according to the P_K and r^2 values.

The statistical analysis showed PK-PD combined models predicting ROR to be less reliable than those predicting LOR. For the three lowest p-values from the two-tailed Fisher exact tests of the patient observations and PK-PD combined model predictions for ROR relative to the 50% and 95% isobols at the mean time, only three of 31 PK-PD combinations were statistically significant. The associations for the other 28 PK-PD combinations were not statistically significant. The averaged p-values and the specific PK-PD combinations are summarized in Table 4 along with best P_K and r^2 values and relevant models.

TABLE 2
OBSERVATIONS AND PHARMACOKINETIC CEFF ESTIMATES AT SURGICAL MILESTONES.

SURGICAL STIMULUS	N	TOTAL OPIOID (NG/ML)	PROP (UG/ML)
Observed Loss of Responsiveness	23	6.83±2.19	1.00±0.91
Laryngoscopy and Tracheal Intubation	13 NR 11 R	6.96±1.86 5.81±1.45	2.66±0.86 2.31±0.64
First Skin Incision	23 NR 1 R	5.90±1.94 4.23	2.82±0.66 1.57
Observed Return of Responsiveness	23	2.83±1.59	1.95±0.42

NR indicates the number of patients who did not respond to pain at the surgical milestone. R indicates the number of patients who responded to pain at the surgical milestone.

TABLE 3
SUMMARY OF THE STATISTICAL TESTS FOR THE COMBINED PK-PD MODELS PREDICTING LOR

TEST	AVERAGE OF TOP TEST SCORES (SD)	BEST PK-PD MODEL COMBINATION	MEAN TEST SCORE FOR ALL PK-PD COMBINATIONS (SD)
Fisher Exact Test, 50% Isobol	p = 0.0051 (0.0078) (statistically significant scores)	*:*:*	P = 0.0051 (0.0078)
Fisher Exact Test, 95% Isobol	p = 0.0014 (0.0040) (statistically significant scores)	*:*:*	P = 0.0014 (0.0040)
P_K	0.790 (0.016) (top 3 scores)	S:*:4,3,2	0.727 (0.032)
r^2	0.8769 (0.0091) (top 3 scores)	TG:1.2:3,2,M	0.6701 (0.1629)

TG, TS, G, and S refer to the Tackley-Gepts, Tackley-Schnider, Gepts, and Schnider propofol PK sets, respectively.
1.2, 1.7, and 2.4 are the three fentanyl relative potencies modeled.
4, 3, 2, and M refer to the sedation PD models of Kern *et al.* for OAA/S < 4, OAA/S < 3, OAA/S < 2, and the Merten's Awakening.
*'s refer to all applicable PK or PD models tested.

TABLE 4
SUMMARY OF THE STATISTICAL TESTS FOR THE COMBINED PK-PD MODELS PREDICTING ROR

TEST	AVERAGE OF TOP 3 TEST SCORES (SD)	BEST PK-PD MODEL COMBINATION	MEAN TEST SCORE FOR ALL PK-PD COMBINATIONS (SD)
Fisher Exact Test, 50% Isobol	p = 0.1910 (0.0186)	TG:1.7:2 TS:1.2,1.7:3 S:1.2:4,M	0.7386 (0.2889)
Fisher Exact Test, 95% Isobol	p =0.0974 (0.0460)	TS:*:2 G:*:3,2	0.4150 (0.2764)
P_K	0.601 (0.002)	TG:1.2:3 TS:2.4:2 G:1.2:3 S:1.2:3,2 S:1.7:3	0.567 (0.025)
r^2	0.1008 (0.0033)	G:1.2:2 S:1.2,1.7:2	0.0366 (0.0366)

TG, TS, G, and S refer to the Tackley-Gepts, Tackley-Schnider, Gepts, and Schnider propofol PK sets, respectively.
1.2, 1.7, and 2.4 are the three fentanyl relative potencies modeled.
4, 3, 2, and M refer to the sedation PD models of Kern *et al.* for OAA/S < 4, OAA/S < 3, OAA/S < 2, and the Merten's Awakening, respectively.

For TI, four PK-PD combined models have a statistically significant relationship with the observations of the patients according to the two-tailed Fisher exact test relative to the 95% isobol. Due to a lack of PK-PD combined model predictions below 50%, the Fisher test could not be used for all PK-PD combinations. However, for the few combinations that were tested relative to the 50% isobol, the predictions are apparently random. According to the P_K and the r^2 tests, the predictions are only slightly better than randomly guessing the anesthetic effect. All the available p-values relative to the 50% isobol, the lowest three p-values relative to the 95% isobol, the three highest P_K-values, the three highest r^2-values, and the specific PK-PD combinations are in Table 5.

For the SI data and using the two-tailed Fisher exact test relative to the 50% and 95% isobols, nine of 21 of the best p-values were statistically significant. Due to high PK-PD predictions, only five model combinations were tested relative to the 50% isobol. However, these high predictions combined with only a single patient response to the first SI resulted in P_K and r^2 average values of 0.989 and 0.9956 for the top 3 scores, respectively. These results and the specific models are shown in Table 6.

TABLE 5
SUMMARY OF THE STATISTICAL TESTS FOR THE COMBINED PK-PD MODELS PREDICTING SI

TEST	AVERAGE OF TOP 3 TEST SCORES (SD)	BEST PK-PD MODEL COMBINATION	MEAN TEST SCORE FOR ALL PK-PD COMBINATIONS (SD)
Fisher Exact Test, 50% Isobol	p = 1.000 (0) {only 1 score available}	TG:*:*	n/a
Fisher Exact Test, 95% Isobol	p = 0.1798 (0.1277)	TG:*:ML TG:2.4:L G:1.2,1.7:MI	0.4216 (0.3164)
P_K	0.748 (0.007)	TG,TS,G:1.2:MI	0.698 (0.033)
r^2	0.1121 (0.0069)	TG:1.2,1.7:ML S:1.2:ML	0.0478 (0.0376)

TG, TS, G, and S refer to the Tackley-Gepts, Tackley-Schnider, Gepts, and Schnider propofol PK sets, respectively.
1.2, 1.7, and 2.4 are the three fentanyl relative potencies modeled.
L, ML, and MI refer to the PD models of Kern *et al.* for Laryngoscopy, and of Mertens *et al.* for Laryngoscopy Alone, and of Mertens *et al.* Tracheal Intubation, respectively.
*'s refer to all applicable PK or PD models tested.

TABLE 6
SUMMARY OF THE STATISTICAL TESTS FOR THE COMBINED PK-PD MODELS PREDICTING TI

TEST	AVERAGE OF TOP 3 TEST SCORES (SD)	BEST PK-PD MODEL COMBINATION	MEAN TEST SCORE FOR ALL PK-PD COMBINATIONS (SD)
Fisher Exact Test, 50% Isobol	P = 0.0667 (0.0228) {Only 2 scores available}	G:1.2:Tet S:1.2:Alg S:*:Tet	n/a
Fisher Exact Test, 95% Isobol	p = 0.0729 (0.0323)	TG:*:Tet TS:*:Tet G:*:Tet G:1.2:Alg S:*:*	n/a
P_K	0.989 (0.008)	TG:2.4:Alg TS:2.4:Alg, Tet	0.853 (0.150)
r^2	0.9956 (0.0033)	G:2.4:Alg TS:1.2:Alg G:*:Alg	0.9105 (0.1119)

TG, TS, G, and S refer to the Tackley-Gepts, Tackley-Schnider, Gepts, and Schnider propofol PK sets, respectively.
1.2, 1.7, and 2.4 are the three fentanyl relative potencies modeled.
Alg and Tet refer to the PD models of Kern *et al.* of shin algometry and electrical tetany, respectively.
*'s refer to all applicable PK or PD models tested.

IV. DISCUSSION

The aim of this study was to observe whether PK-PD model predictions were accurate predictors of anesthesia in the operating room at four clinical milestones. We also hoped this study would help relate surrogate measures of pain to surgical stimuli. However, the lack of measured drug concentrations, exasperated by a lack of control over individual dosing schemes, make it impossible to distinguish between pharmacokinetic and pharmacodynamic variances. The pharmacodynamic relationship between the surgical stimuli and surrogate measures remains unclear.

Though we observed patient responses to surgical stimulation throughout the entire surgery of each patient, we were unprepared for the subtle differences between cholecystectomies, hernia repairs, and nissen fundoplications. The four clinical milestones of LOR, ROR, TI, and SI were chosen because they were consistently identifiable for all these TIVA-appropriate surgeries. Had we observed more patients for the same types of surgeries, we would expect to have reported on the PK-PD combined model predictions for other specific surgical stimuli.

A fundamental challenge for this study was the degrees for freedom we allowed while considering numerous variables. For example, a key principle in creating and validating drug interaction models is to exercise control over the drug concentrations. It was overly optimistic to forgo this control—no plasma samples were taken and the anesthetists were only asked to follow their (individual) standard practices to provide total intravenous anesthesia using propofol and remifentanil as the primary anesthetic agents.

We must also remember that although sedation was treated as a binary state in the operating room, it is actually a continuous measure. This discrepancy is compounded by our inability to identifying the exact moments of LOR and ROR through some repeated stimulus requiring a single clear

response. For example, in a controlled study environment, a volunteer may be queried 3 to 6 times a minute to observe the exact moment of LOR. In our observational study, the moment of LOR was assessed by a research nurse watching the patient and the anesthesiologist but without addressing the patient.

Because this study was only observational, tracheal intubation followed laryngoscopy as quickly as possible. As a result, we were unable to separate these two milestones, making it difficult to observe the potential differences in predicting the anesthesia necessary for these two noxious stimuli.

The milestone of SI also presented a surprising dilemma—although we observed the first skin incision clearly, the second, third, fourth, *etc*. incisions were less obvious. We wished to use all the data we had gathered from these observations, yet lack the statistical tools to know how to evaluate repeated stimuli in the same patient—this violates a fundamental assumption of independence, necessary for most statistical tests.

Despite clear criteria to identify patient responses to pain, pain itself is a subjective measure. Furthermore, the definition of "adequate anesthesia" is inexact. For example, the anesthesiologist may respond to a 15% rise in heart rate by increasing the opioid. Since we defined a heart rate response to pain to be a rise exceeding 20%, this episode was not considered a response to a noxious stimulus. However, was "adequate anesthesia" provided seeing that the patient began to respond? What would have been the ideal delivery of anesthesia, to prevent all responses to noxious stimuli, or to titrate the dosing in order to minimize the response itself? In our study, we treated rising heart rates that were then controlled, as nonresponses to pain. Our PD models were simply not designed to describe the continuous or graded nature of pain for an individual. Instead, we assume this continuous scale to be somehow described by the population variance of a binary assessment of "adequate anesthesia."

Some may question the different methods of used to identify the most predictive PK-PD combined model. We found a dearth of statistical methods designed specifically to validate response surface data in a monitoring environment. We were disappointed that the Fisher exact test essentially reduces response surfaces into isobologram data. It is also somewhat unwieldy, requiring the manual calculation of contingency tables relative to each isobol of interest. Additionally, the P_K-test is not yet well understood, and appears to test the rank-order of the data instead of directly evaluating the accuracy of the predictions themselves. The r^2-value fails to account for the spread of concentrations of the response surface data, but only compares the final PD prediction with the observations. It was our hope that this combination of statistical tests identified PK-PD combined models that have a positive predictive value.

We recognize that this study does not provide a clear "gold standard" based on globally optimized data—we could have attempted to create a single overall PK-PD model by changing each single variable until we found a single optimal model. We were not so ambitious, realizing that with another data set, each PK and PD variable would change. Such a study would also ignore the findings of previous research, and would thus prove inadaptable to new research. This study hoped to take advantage of the best results from multiple studies to find a PK-PD combined model that works well.

Future protocols would outline a slower induction by infusion, in order to minimize the differences between bolus and infusion pharmacokinetics and dynamics. This slower induction would also increase the accuracy of the pharmacokinetic predictions by reducing the peak and drop-off of bolus kinetics that is particularly hard to predict. Though plasma samples would help validate pharmacokinetic models, it is clear that the estimated Ceff would still need to be related to the observed pharmacodynamics. An automated system that requires a response from the patient may provide a more consistent and precise measure of LOR and ROR. Though hazarding statistical independence of measures, we would consider the use of repeated surgical stimuli, while titrating the dosing. Titration throughout the surgery may result in a better evaluation of intraoperative PD models, such as for SI. A similar scheme was successfully used by Mertens *et al.* to create a laryngoscopy response surface directly from patient data.

Tracheal intubation is generally attempted when Ceffs are expected to be peaking. Surprisingly, our data showed that laryngoscopy and tracheal intubation often occur after the concentrations had peaked and begun to stabilize.

The large number of LOR PK-PD combined models providing accurate predictions compared to the fewer number of ROR PK-PD combined models is slightly surprising. This difference may be due to PD hysteresis between LOR and ROR; it is recognized that both PK and PD change with the duration of surgery. Due to a lack of assayed anesthetic concentrations, we are unable to confirm whether this was a significant factor in the relatively short surgeries that we observed.

We believe another significant difference is the nature of the dosing at induction verses awakening—following a bolus (such as given at induction) the kinetics rise and fall rapidly while awakening at the end of surgery is at relatively steady-state concentrations as the anesthetics are eliminated from the body. Not only are bolus PK difficult to estimate, but PD responses are different depending on the dosing scheme.

For TI and SI, the PK-PD combined model predictions were high enough that a contingency table relative to the 50% and 95% isobols had columns filled with 0's. In other words, most of these predicted data points were on the "flat" portion of the response surface. According to these model predictions, the patients would have been sufficiently anesthetized for TI and SI at lower dosing schemes. Due to relatively high anesthetic concentrations at TI and SI, it is difficult to identify which PD model best describes the anesthesia necessary to prevent a

response to these noxious stimuli. This most clearly summarizes the challenge of an observational study—the conservative practice of clinicians' results in very few responses to pain in patients, especially at early stages of the surgery. A study titrating the anesthetic early on or by observing surgical stimuli later in the surgery might alleviate this challenge. Merten *et al.* used a predefined protocol to titrate the drug in order to gather data both when anesthesia was adequate and inadequate. However, repeated measures face the statistical challenge of losing their independence. This remains a statistical conundrum in anesthesiology research.

V. Conclusion

The purpose of this experiment was to study the predictive value of combined PK-PD models. The most frequent PK models for the best predictions of adequate anesthesia for all clinical stimuli are the Schnider propofol model and a fentanyl scaling factor of 1.2. For the PD models, the OAA/S<2 model is the best for sedation (both LOR and ROR). The laryngoscopy and the tracheal intubation models of Mertens *et al.* are equally predictive for TI and the electrical tetany model appears to be best for SI. This observational study will hopefully serve as a preliminary step in applying PK and PD modeling clinically.

References

[1] Minto CF, Schnider TW, Egan TD, Youngs E, Lemmens HJ, Gambus PL, Billard V, Hoke JF, Moore KH, Hermann DJ, Muir KT, Mandema JW, Shafer SL: Influence of age and gender on the pharmacokinetics and pharmacodynamics of remifentanil. I. Model development. Anesthesiology 1997; 86 (1): 10-23

[2] Shafer SL, Varvel JR, Aziz N, Scott JC: Pharmacokinetics of fentanyl administered by computer-controlled infusion pump. Anesthesiology 1990; 73 (6): 1091-1102

[3] Minto CF, Schnider TW, Gregg KM, Henthorn TK, Shafer SL: Using the Time of Maximum Effect Site Concentration to Combine Pharmacokinetics and Pharmacodynamics. Anesthesiology 99 (2): 324-333

[4] Scott JC, Stanski DR: Decreased fentanyl and alfentanil dose requirements with age. A simultaneous pharmacokinetic and pharmacodynamic evaluation. J Pharmacol Exp Ther. 1987 Jan; 240 (1): 159-66

[5] Tackley RM, Lewis GTR, Prys-Roberts C, Boaden RW, Dixon J, Harvey JT: Computer Controlled Infusion of Propofol. Br J Anaesth 1989; 62: 46-53

[6] Kern SE, Egan TE, White JL, Cluff M: Characterizing pharmacodynamic synergism between propofol and remifentanil in volunteers using response surfaces. Anesthesiology 1999; 91: A342

[7] Gepts E, Camu F, Cockshott ID, Douglas EJ: Disposition of Propofol administered as constant rate intravenous infusions in humans. Anesth Analg 1987; 66 (12): 1256-63

[8] Schnider TW, Minto CF, Shafer SL, Gambus PL, Andresen C, Goodale DB, Youngs EJ: The Influence of Age on Propofol Pharmacodynamics. Anesthesiology 1999; 90 (6): 1502-16

[9] Egan TD, Muir KT, Hermann DJ, Stanski DR, Shafer SL: The electroencephalogram (EEG) and clinical measures of opioid potency: defining the EEG-clinical potency relationship ('fingerprint') with application to remifentanil. International Journal of Pharmaceutical Medicine 2001; 15: 1-9

[10] Chernik DA, Gillings D, Laine H, Hendler J, Silver JM, Davidson AB, Schwam EM, Siegel JL: Validity and Reliability of the Observer's Assessment of Alertness/Sedation Scale: Study with Intravenous Midazolam. J of Clin Psychopharmacol 1990; 10 (4): 244-251

[11] Mertens MJ, Olofsen E, Engbers FHM, Burm AGL, Bovill JG, Vuyk J: Propofol Reduces Perioperative Remifentanil Requirements in a Synergistic Manner: Response Surface Modeling of Perioperative Remifentanil-Propofol Interactions. Anesthesiology 2003; 99 (2): 347-359

[12] Glantz, S. A., *Primer of Biostatistics*. USA: McGraw-Hill, 2002, pp. 264-273.

[13] Smith, W. D., Dutton, R. C., and Smith, N. T., "Measuring the performance of anesthetic depth indicators," Anesthesiology, Vol. 84, pp. 38-51, 1996

[14] Smith, W. D., Dutton, R. C., and Smith, N. T., "A measure of association for assessing prediction accuracy that is a generalization of non-parametric ROC area," Statistics in Medicine, Vol. 15, pp. 1199-1215, 1996