Response Surface Model Predictions of Wake-Up Time During Scoliosis Surgery

Chien-Kun Ting, MD, PhD,* Ken B. Johnson, MD,† Wei-Nung Teng, MD,* Noah D. Synoid, MS,† Cris LaPierre, PhD,§ Lu Yu, MS,¶ and Dwayne R. Westenskow, PhD††

BACKGROUND: With the use of previously published data, new sevoﬂurane–remifentanil interaction models of various degrees of sedation were created and adapted to desﬂurane–fentanyl by using minimal alveolar concentration and opioid equivalencies. These models were used to predict return of responsiveness in patients undergoing scoliosis surgery during a wake-up test. Our hypothesis was that one of the interaction models would accurately predict return of responsiveness during a wake-up test.

METHODS: Three new sevoﬂurane–remifentanil interaction models were constructed from previous observations in volunteers by using the Observer’s Assessment of Alertness/Sedation (OAA/S) scores. These models included predictions of OAA/S<2 (unresponsive), OAA/S<3, and OAA/S<4 (sedation). Twenty-three patients scheduled for scoliosis surgery received a fentanyl–desflurane anesthetic. With the use of published pharmacokinetic models, predictions of fentanyl and desflurane effect-site concentrations were recorded throughout surgery and converted to equivalent remifentanil and sevoﬂurane effect-site concentrations. Data were recorded every 30 seconds from the time when desflurane was turned off until 10 minutes after the patients responded by moving their hands and toes. Model predictions were compared with observations with graphical and temporal analyses.

RESULTS: The average difference between the time when a patient first responded and the time when the model predicted that there was a 50% probability that the patient would respond were −2.6 ± 3.6 minutes (mean ± SD) for the OAA/S<2 model, 2.8 ± 5.6 minutes for the OAA/S<3 model and 52.6 ± 32.3 minutes for the OAA/S<4 model.

CONCLUSIONS: The results confirmed our study hypothesis; a sevoﬂurane–remifen-tanil interaction model built from observations in volunteers and adapted to desflurane and fentanyl accurately predicted patient response during a wake-up test. These results were similar to our previous study comparing model predictions and patient observations after a sevoﬂurane–remifentanil/fentanyl anesthetic. The OAA/S<2 model most accurately predicted the time patients would respond by moving their fingers and toes. This model may help anesthesiologists better predict return of responsiveness during a wake-up test in patients undergoing spine surgery. (Anesth Analg 2014;118:546–53)

Postoperative neurologic deﬁcit is one of the most feared complications of complex spinal reconstructive surgery.¹ The Stagnara wake-up test has been used to assess neurologic deﬁcits during scoliosis surgery since 1973.² Many anesthesiologists now prefer to monitor motor- and sensory-evoked potentials; however, the wake-up test is still used by many surgeons.³−⁴ The wake-up test is a unique challenge for an anesthesiologist; allow the patient to briefly emerge from anesthesia to participate in a neurologic assessment, yet remain adequately sedated to tolerate incision pain and an endotracheal tube.⁵

Response surface interaction models predict anesthetic drug effect (e.g., sedation, analgesia, etc). They may be useful to anesthesiologists in dosing opioids, sedatives, and potent inhaled drugs to optimize sedation and analgesia for a wake-up test.⁶−⁸ However, previous published interaction models for loss of responsiveness may not adequately predict when patients will emerge from anesthesia to participate in a wake-up test. The principle aim of our study was to develop a new interaction model for opioids and inhaled sevoﬂurane in volunteers that predicts the probability of
loss of response to auditory and tactile stimuli by using the Observer’s Assessment of Alertness/Sedation (OAA/S) score presented in Table 1.13 Volunteers were considered unresponsive if the OAA/S score was 0 or 1. Volunteers were considered responsive if the OAA/S score was higher than or equal to 2. The threshold of OAA/S<2 was used to distinguish between responsive and unresponsive.11,12 Their model used a modified Greco construct equation 1.14,15

\[
\text{Effect} = \frac{E_{\text{max}} \times \left[ \frac{C_{r}}{C_{r}^{50}} + \frac{C_{s}}{C_{s}^{50}} - \left( \frac{C_{r}}{C_{r}^{50}} \times \frac{C_{s}}{C_{s}^{50}} \right) \right]^n + 1}{C_{r}^{50} + C_{s}^{50}}
\]

(1)

Effect ranged from 0 (100% probability of response) to 1 (100% probability of no response) where \(E_{\text{max}}\) is fixed as the maximal effect (100% probability of no response), and when the remifentanil and sevoflurane concentrations were 0, the probability of effect is set to 0. \(C_{50r}\) (sevoflurane) and \(C_{50r}\) (remifentanil) are the effect-site concentrations that produce 50% of the maximal effect when sevoflurane or remifentanil is administered individually, \(n\) is the steepness of the response surface, and \(\alpha\) is the interaction parameter between sevoflurane and remifentanil.

We used a wide range of remifentanil and sevoflurane effect-site concentrations and OAA/S scores to create 3 new models of sedation.10 We used modeling software (Matlab, The Mathworks, Natick, MA) and a naïve pooled technique to develop a model to predict the drug concentrations at which patients would be unresponsive to moderate prodding or shaking (OAA/S<2), a model to predict when patients would respond to moderate prodding or shaking that we defined as moderate sedation (OAA/S<3) and a model to predict when patients would respond to their name being called loudly that we defined as minimal sedation (OAA/S<4). Model parameters were determined by using an iterative approach with Matlab routine “fminsearch” to find the minimum value for –2LL in equation 2:

\[
-2LL = -2 \sum_{i=1}^{N} \left[ Ri \times \text{Ln}(P) + (1 - Ri) \times \text{Ln}(1 - P) \right],
\]

(2)

\(N\) is the number of observations made for all volunteers combined, \(R\) is the observed response, and \(P\) is the corresponding probability of loss of response.

Model parameters are listed in Table 2.

To assess the fit of each new model to observed responses in volunteers, we calculated the percentage of each of the model’s predictions of responsiveness that matched the corresponding observed responses in volunteers. The model prediction and the observation were considered to agree with each other when the absolute difference between the model predicted probability of a certain OAA/S score and the observed OAA/S score was <0.5. For example, if the model predicted that there was a 0.6 probability that the OAA/S score was <3 (Probability of OAA/S<3 = 0.6), and the observation yielded an OAA/S score of 2 (OAA/S = 2; therefore, probability of OAA/S<3 = 1.0), then 1.0 – 0.6 = 0.4, which is <0.5, the prediction was considered to match the observation. Responses in volunteers were compared with model predictions by using a Spearman rank correlation (\(\rho\)). A 2-tailed unpaired \(t\) test was used to determine whether the Spearman \(\rho\) was significantly different from 0. The null hypothesis was that the new models did not correlate with observations.

### Evaluation of Response Surface Models in Patients Undergoing Scoliosis Surgery with Wake-Up Test

After IRB approval at the Taipei Veterans General Hospital, the requirement for written informed consent was waived by the IRB. We reviewed the medical and anesthesia records of all ASA physical status class I to II patients who had surgery for the correction of scoliosis under general anesthesia with desflurane and fentanyl between 2005 and 2011. We
excluded patient records with missing demographic data, neurological disability, impaired hearing, a history of epilepsy or ongoing opioid consumption and those taking anti-epileptic, sedative, or stimulant medication. We found that the records from 23 patients (8 men and 15 women) met all criteria for inclusion.

Review of the anesthesia records revealed that a standard anesthesia protocol had been followed for each patient: induction with propofol (2 mg/kg) and fentanyl (5 μg/kg), cisatracurium, or rocuronium to facilitate endotracheal intubation, maintenance with fentanyl and desflurane (titrated at the anesthesiologist’s discretion), and intermittent IV bolus injections of cisatracurium to maintain 1 twitch after a train-of-four stimulus. The tidal volume was set at 8 mL/kg, and the respiratory rate was adjusted to keep the end-tidal carbon dioxide (ETCO₂) between 32 and 36 mm-Hg. Following the standard protocol, the administration of cisatracurium was interrupted 30 minutes before the anticipated start of the wake-up test. Small doses of edrophonium and atropine were given if the train-of-four count did not recover to four-of-four twitches, with no fade. The fentanyl infusions were terminated 14.5 ± 9.2 minutes (mean ± SD) before the start of the wake-up test. The desflurane vaporizer was turned off, and the fresh gas flow was increased from 2 to 4 L/min just before the wake-up test. At the start of the wake-up test, patients were called by their first name in a loud fashion while moderately moving their shoulder and asking, “Can you move your fingers and toes?” If movement did not occur within 2 seconds, the question was repeated by using a louder voice. The questioning was repeated every 15 seconds until the patient responded. Wake-up time was defined as the elapsed time between the time when the vaporizer was turned off and the time when the patient first moved their fingers and toes.

Model Predictions of Responsiveness
A pharmacokinetic model was used to calculate fentanyl effect-site concentrations throughout the procedure by using the time and size of each administered fentanyl dose. Fentanyl effect-site concentrations were converted to remifentanil equivalents by using a remifentanil to fentanyl potency ratio of 1:1.2. A modified multi-compartment pharmacokinetic model was used to calculate desflurane brain tissue concentrations throughout the procedures from the end-tidal desflurane concentrations recorded throughout the procedure. Desflurane brain tissue concentrations were converted to sevoflurane equivalents by using a sevoflurane to desflurane potency ratio of 3.3:1. The 3 sevoflurane–remifentanil response surface models for OAA/S<2, OAA/S<3, and OAA/S<4 (equation 1, Table 2) were used to calculate the probability that the patients would or would not respond with or without moderate prodding or shaking and calling their name loudly. Model predictions were calculated every 10 seconds from the time desflurane was turned off until 10 minutes after the patients responded by moving their hands and toes.

Statistics
The time difference between the time when each patient first responded by moving his/her fingers and toes and the time when each of the 3 OAA/S models predicted that there was a 50% probability that the patient would respond were assessed by 1-way analysis of variance. If significant, a post hoc test with Bonferroni correction to account for multiple comparisons was performed to identify differences between individual models. Results were considered significant with a P value <0.05. Statistical analyses were performed with SPSS version 17.0 for Windows (SPSS Inc., Chicago, IL).

RESULTS
Response Surface Model Development
Table 2 lists the best-fit parameters and their respective coefficients of variation for the 3 sevoflurane–remifentanil interaction models: no response to moderate prodding or shaking (OAA/S<2), response to moderate prodding or shaking (OAA/S<3), and response when name is called loudly (OAA/S<4).

With the use of a probability of 0.5 or higher as a cut-off for model goodness of fit, Table 2 lists the percentage of model predictions of responsiveness consistent with observed responses in volunteers. The Spearman rank correlation coefficients and the 2-tailed t test analysis rejected the null hypothesis that the revised models did not correlate with observations in volunteers.

Evaluation of Response Surface Models in Patients Undergoing Scoliosis Surgery
Table 3 lists the demographics and surgical times for the 23 patients in the study. Three surgeons and 11 anesthesiologists participated in the study. After recovery, none of the patients reported recalling any portion of the procedure except for events that occurred during the wake-up test.

Figure 1 shows the model predictions of the probability that each patient would not respond to moderate prodding or shaking (OAA/S<2), would respond to moderate prodding or shaking (OAA/S<3), and would respond to their name being called loudly (OAA/S<4) during the wake-up test. The vertical lines show the time when the patient responded by moving their fingers and toes. The average time between when the vaporizer was turned off to when patients moved their fingers and toes was 10.3 ± 4.7 minutes (mean ± SD).

When each of the patients first responded by moving their fingers and toes, we used the pharmacokinetic models to predict the drug effect-site concentrations at the time they woke up. We used these effect-site concentrations and the 3 OAA/S pharmacodynamic response surface models to calculate the probability that the patient would or would not

| Age(y) | 15.9 ± 2.9 |
| No. male | 8 |
| No. female | 15 |
| No. ASA physical status I | 13 |
| No. ASA physical status II | 10 |
| Weight (kg) | 51.2 ± 18.6 |
| Height(cm) | 162.0 ± 6.6 |
| Body mass index (kg/m²) | 19.5 ± 3.1 |
| Time from intubation to end of wake-up test (min) | 278.0 ± 64.9 |
| Time after wake-up test to end of anesthesia (min) | 162.3 ± 31.1 |
Figure 1. Model predictions of the Observer's Assessment of Alertness/Sedation (OAA/S) score during the wake-up test for each patient. The dashed line shows the probability that a patient would respond to their name spoken in a normal tone (OAA/S<4), the dashed line represents the probability that a patient would respond to their name called loudly (OAA/S<3), and the solid line shows the probability that a patient would respond only after moderate prodding or shaking (OAA/S<2). The gray vertical line shows the time when each patient first moved their fingers and toes after their first name was called, having their shoulders moved moderately and being asked, “Can you move your fingers and toes?”
respond to stimuli at the time they woke up: no response to moderate prodding or shaking (OAA/S<2), response to moderate prodding or shaking (OAA/S<3), and response to their name being called loudly (OAA/S<4). Figure 2 shows the model predictions of probability of response (for each of the 3 OAA/S models) versus the percentage of patients who responded to each of the stimuli. The 1-way analysis of variance indicated that there was a difference in emergence time predictions between models ($P < 0.001$). The post hoc analysis revealed that predictions by using the OAA/S<2 model (unresponsiveness) and the OAA/S<3 model (moderate sedation) better matched observations than predictions by using the OAA/S<4 model (minimal sedation) ($P < 0.001$), but there was no significant statistical difference between them ($P = 0.946$).

Figure 3 shows the difference between the time when each patient first responded by moving his/her fingers and toes and the time when each of the 3 OAA/S models predicted that there was a 50% probability that the patient would respond. If the model was 100% accurate, all differences would be zero. The average difference and the standard deviation (SD) of the differences were $-2.6 \pm 3.6$ minutes for the OAA/S<2 model (no response to prodding and shaking), $2.8 \pm 5.6$ minutes for the OAA/S<3 model (response to prodding and shaking), and $52.6 \pm 32.3$ minutes for the OAA/S<4 model (response to name called loudly).
DISCUSSION

Our results showed that the model of OAA/S<2 had a bias that predicted emergence just before (2–3 minutes) observed emergence. However, the model of OAA/S<3 had a bias that predicted emergence just after (also 2–3 minutes) observed emergence. The model of OAA/S<4 had a larger bias that predicted emergence long after observed emergence. Thus, the 2 models have similar absolute errors and from a statistical analysis standpoint are very similar. However, in terms of clinical utility, we propose that predictions of emergence before the actual emergence are most helpful in the setting of a wake-up test. Hence, the OAA/S<2 model worked best. This response surface model may help anesthesiologists prepare for a wake-up test if the model is used in real time to predict the time remaining until the patient will wake up, as a function of previously delivered doses of opioids and inhaled drugs.

Managing drug delivery during the wake-up test is difficult. The anesthesiologist is expected to manage the patient’s anesthetic state such that the patient briefly emerges from anesthesia to participate in a neurologic assessment yet remains adequately sedated to tolerate incision pain and an endotracheal tube. During the wake-up test, once the desflurane vaporizer was shut off, the difference between model predictions of emergence, defined as a 50% or more probability of response, and when patients actually woke up to participate in a neurologic assessment was 2.6 ± 3.6 minutes (mean ± SD). Given that the average time to wake up was 10.3 ± 4.7 minutes after starting the wake-up test, the accuracy of the model’s prediction of wake-up time seems sufficient to be of clinical value.

Other research groups have found clinical value in applying pharmacodynamic models to predict wake-up time. The Fresenius-Kabi syringe pump (Fresenius-Kabi AG, Bad Homburg, Germany) predicts the time to wake up after the termination of a propofol infusion. The Drager Smart Pilot (Drager Medical, Lubeck, Germany) and the GE Navigator (GE Healthcare, Helsinki, Finland) predict the time to wake up after terminating drug infusions and/or terminating the delivery of inhaled drugs. Experienced users find these devices helpful in managing drug delivery to control wake-up time. The Smart Pilot and the Navigator also include models for response to noxious stimuli that were implemented to help manage analgesia after wake up, but the value of these models has not been evaluated. Likewise, we do not have data to evaluate the value of our models for managing analgesia during the wake-up test.

To our knowledge, ours is the first remifentanil–sevoflurane interaction model that calculates the probability of response to moderate prodding and shaking (OAA/S<3) and the first model to calculate the probability of response to name called loudly (OAA/S<4). These 2 new models fit responses from volunteers reasonably well when we compared the model prediction of probability of response to the number of patients who actually responded; 88% to 90% of the model predictions matched the responses in volunteers. Each model accurately predicted the transition from responsive to unresponsive, as confirmed by the percentage of model predictions consistent with observed responses.

In model construction, by using data from a volunteer study is an advantage by affording the opportunity to study the entire surface of the drug interaction. With the use of volunteer data, we were able to study volunteer responses to noxious stimuli when they received remifentanil alone and when they received a combination of remifentanil and sevoflurane. In patient studies, the drug doses and resulting concentrations are obviously constrained by what is clinically prudent. This is important because it is difficult to characterize the interaction surfaces unless each drug is studied to the point of near maximal effect, in isolation.

Our results indicate that the drug concentrations at which volunteers no longer responded to moderate prodding or shaking are the same concentrations at which scoliosis patients responded after their first name was called loudly and their shoulders were moderately moved while being asked, “Can you move your fingers and toes?” (Figs. 1, 2 and 3). This seems reasonable because, in the volunteer study, the OAA/S assessments were done in the absence of pain and without an endotracheal tube in place. It appears that surgical pain replaces the need to shake the patient’s shoulder to arouse him/her enough to respond to the voice command during the wake-up test.

The model performed as expected; observed responses were distributed about the model’s 50% probability of response. Nearly half of the patients emerged at model predictions below and others above the 50% probability of response. However, because of the small sample size, it is unclear how well model predictions discriminate between the patients’ OAA/S<2 and OAA/S>1 states. With a larger data sample, other statistical methods might have provided a better estimate of model discrimination and calibration.

The Fresenius-Kabi syringe pump (Fresenius-Kabi AG, Bad Homburg, Germany) uses simple propofol pharmacodynamic models to predict the time to wake up after the termination of a propofol infusion. This simple 1-drug model does not consider the effects of opioids given before the wake-up time. The Drager Smart Pilot (Drager Medical, Lubeck, Germany) and the GE Navigator (GE Healthcare, Helsinki, Finland) both use 2 pharmacokinetic models, 1 for propofol (or an inhaled drug) and 1 for the opioid infusion to predict drug concentrations. They then use an interactive pharmacodynamic model to predict time to wake up. Experienced users discuss their potential to manage drug delivery to control wake-up time. They also help manage analgesia after wake up, but this feature has not been evaluated.

There are several limitations to this line of investigation. One limitation of this study was a function of selecting the appropriate volatile drug potency ratios to adapt the sevoflurane portion of the response surface model to desflurane. There are several studies that report the minimal alveolar concentration (MAC) and MACawake values over a range of patient ages. For example, in the age range of 30 to 50 years, MAC ranges from 1.58% to 2.05% for sevoflurane and from 2.42% to 2.6% for desflurane, resulting in a potency range for sevoflurane/desflurane of 1.3 to 1.8. MACawake ranges from 0.61% to 0.70% for sevoflurane and from 2.42% to 2.6% for desflurane, resulting in a potency range of 2.9 to 4.6. For purposes of this analysis, we chose a potency ratio of 3.3, which overlapped the potency ranges based on both MAC and MACawake. A different choice of potency ratio may alter the study’s results.

A second limitation is that a volunteer study cannot fully emulate the complexities of the clinical environment. The differences between the volunteer study on which the models...
Response Surface Model Predictions of Wake-Up Time

are based and the clinical environment in which the models are applied likely impact model performance to some degree. Although we used the modified Greco model to describe the sevoflurane–remifentanil interaction on the OAA/S scale, there are several other available interaction model structures (i.e., logit, hierarchical, etc). Although beyond the scope of this study, the clinical implications of using different interaction model structure on predictions may warrant further investigation.

A third limitation is that all 3 OAA/S models were developed from volunteers who were 24.1 ± 3.5 years old. We applied the models to our patients whose ages ranged from 12 to 24 years with a median of 15 years old under an assumption that the younger-aged patients would behave in a similar fashion as the older volunteers. From the present study, it seems to work well. Further investigations of age limitation for the adult model can be applied should they be studied and justified.

Finally, the residual amount of propofol and midazolam remaining after premedication and anesthetic induction may have been too low to play a role in confounding the result of model predictions and likely had minimal contribution at the end of the anesthetic.11

The model predicting the probability that study volunteers would not respond to moderate prodding or shaking produced the most accurate prediction of the time when patients responded (by moving their fingers and toes) during a wake-up test. This model may help anesthesiologists prepare for a wake-up test by predicting the time when a patient will wake up and may help anesthesiologists manage drug doses when preparing for a wake-up test. ⊂

DISCLOSURES

Name: Chien-Kun Ting, MD, PhD. Contribution: This author helped design and conduct the study, analyze the data, and write the manuscript. Attestation: Chien-Kun Ting has seen the original study data, reviewed the analysis of the data, approved the final manuscript, and is the author responsible for archiving the study files. Conflicts of Interest: The author has no conflicts of interest to declare. Name: Ken B. Johnson, MD. Contribution: This author helped analyze the data and write the manuscript. Attestation: Ken B. Johnson reviewed the analysis of the data and approved the final manuscript. Conflicts of Interest: Ken B. Johnson has equity interest in Applied Medical Visualizations. Ken Johnson is an equity partner in Applied Medical Visualizations, a company that builds “drug display” technology by using response surface model predictions. Name: Wei-Nung Teng, MD. Contribution: This author helped conduct the study. Attestation: Wei-Nung Teng has seen the original study data and approved the final manuscript. Conflicts of Interest: The author has no conflicts of interest to declare. Name: Noah D. Synoid, MS. Contribution: This author helped analyze the data. Attestation: Noah D. Synoid reviewed the analysis of the data and approved the final manuscript. Conflicts of Interest: Noah D. Synoid has equity interest in Applied Medical Visualizations.

Name: Cris LaPierre, PhD. Contribution: This author helped conduct the study and write the manuscript. Attestation: Cris LaPierre reviewed the analysis of the data and approved the final manuscript. Conflicts of Interest: The author has no conflicts of interest to declare. Name: Lu Yu, MS. Contribution: This author helped analyze the data and write the manuscript. Attestation: Lu Yu reviewed the analysis of the data and approved the final manuscript. Conflicts of Interest: The author has no conflicts of interest to declare. Name: Dwayne R. Westenskow, PhD. Contribution: This author helped design and conduct the study and write the manuscript. Attestation: Dwayne R. Westenskow reviewed the analysis of the data and approved the final manuscript. Conflicts of Interest: The author has no conflicts of interest to declare. This manuscript was handled by: Tony Gin, FANZCA, FRCA, MD.

REFERENCES

12. Syroid ND, Johnson KB, Pace NL, Westenskow DR, Tyler D, Brühschwein F, Albert RW, Roalstad S, Costy-Bennett S, Egan TD. Response surface model predictions of emergence


