# **Predictive Performance of Three Multivariate Difficult Tracheal Intubation Models: A Double-Blind, Case-Controlled Study**

Mohamed Naguib, MB, BCh, MSc, FFARCSI, MD\*, Franklin L. Scamman, MD‡, Cormac O'Sullivan, CRNA‡, John Aker, CRNA§, Alan F. Ross, MD‡, Steven Kosmach, MSN, RN\*, and Joe E. Ensor, PhD+

Departments of \*Anesthesiology and Pain Medicine and †Biostatistics and Applied Mathematics, The University of Texas M. D. Anderson Cancer Center, Houston; ‡Department of Anesthesia, The University of Iowa Roy J. and Lucille A. Carver College of Medicine, Iowa City; and §Department of Anesthesia, Children's Mercy Hospitals & Clinics, Kansas City, Missouri

We performed a case-controlled, double-blind study to examine the performance of three multivariate clinical models (Wilson, Arné, and Naguib models) in the prediction of unanticipated difficult intubation. The study group consisted of 97 patients in whom an unanticipated difficult intubation had occurred. For each difficult intubation patient, a matched control patient was selected in whom tracheal intubation had been easily accomplished. Postoperatively, a blinded investigator evaluated both patients. The clinical assessment included the patient's weight, height, age, Mallampati score, interincisor gap, thyromental distance, thyrosternal distance, neck circumference, Wilson risk sum score, history of previous difficult intubation, and diseases associated with difficult laryngoscopy or intubation. The Naguib model was

significantly more sensitive (81.4%; P < 0.0001) than the Arné (54.6%) or Wilson (40.2%) models. Both the Naguib (76.8%) and Arné (74.7%) model classified more intubations correctly (P = 0.01) than the Wilson model (66.5%). The specificity of Arné, Wilson, and Naguib model was 94.9%, 92.8%, and 72.2%, respectively (P < 0.0001). The corresponding area under the receiver operating characteristic curve was 0.87, 0.79, and 0.82, respectively. Our new model for prediction of difficult intubation was developed using logistic regression and includes thyromental distance, Mallampati score, interincisor gap, and height. This model is 82.5% sensitive and 85.6% specific with an area under the receiver operating characteristic curve of 0.90.

(Anesth Analg 2006;102:818-24)

Inanticipated difficult tracheal intubation is a significant source of morbidity and mortality in anesthetic practice. The incidence of difficult intubation in the operating room has been reported to range from 1% to 18% (1–4). The incidence of abandoned/failed intubation is approximately 0.05%– 0.35% (5,6), whereas that of cannot ventilate by mask, cannot intubate is around 0.0001%–0.02% (7). In the Confidential Enquiry into Maternal Deaths in Great

DOI: 10.1213/01.ane.0000196507.19771.b2

Britain for 2000–2002, 3 of the 6 deaths directly attributable to anesthesia were associated with failed tracheal intubation (8). Worldwide, up to 600 patients are thought to die annually as a result of complications occurring at the time of tracheal intubation (9). Approximately 30% of the deaths in patients who experienced difficulties at laryngoscopy or intubation are caused by hypoxic brain damage secondary to inability to maintain a patent airway (2).

Difficult tracheal intubation accounted for approximately 17% of adverse respiratory events in an American Society of Anesthesiologists closed-claims analysis (10). In 85% of these cases, the outcome was either death or brain damage (10). Increases in the incidence of morbid nonfatal events have also been noted in patients who have undergone difficult tracheal intubation (11–14). These events included desaturation, hypertension, esophageal intubation, pharyngeal trauma, dental injury, cancellation of surgery, increased hospital stay, and

Presented, in part, at the American Society of Anesthesiologists Annual Meeting, October 22–26, 2005, New Orleans, LA.

Accepted for publication October 17, 2005.

Address Correspondence and reprint requests to Mohamed Naguib, MB, BCh, MSc, FFARCSI, MD, Department of Anesthesiology and Pain Medicine, Unit 409, The University of Texas M. D. Anderson Cancer Center, 1400 Holcombe Blvd., Houston, TX 77030. Address e-mail to Naguib@mdanderson.org.

an increased rate of unexpected intensive care unit admission.

In most studies, difficult laryngoscopy has been defined as a view of the larynx corresponding to grade 3 or 4 in the classification of difficult intubation by Cormack and Lehane (5). The American Society of Anesthesiologists defines difficult tracheal intubation as when "proper insertion of the endotracheal tube with conventional laryngoscopy requires more than 3 attempts, or more than 10 min" (15).

Although unanticipated difficult intubation has been the subject of many studies, a puzzling feature of these studies is the wide variation in the reported sensitivity of the different models used for prediction of this problem (1,3,4,6,16–22). A test performed to predict difficult intubation should have high sensitivity so that it will identify most patients in whom intubation will truly be difficult. We are not aware of any studies that have evaluated different multivariate models in the same population of patients to determine the most sensitive model for predicting difficult intubation. Therefore, we designed and performed the double-blind, case-controlled study described herein to compare and validate the predictive performance of three multivariate clinical models described by Wilson et al. (1), Arné et al. (20), and Naguib et al. (21) in a group of patients who had confirmed unanticipated difficult intubation. The sensitivity reported for the latter two models has been the highest sensitivity for such clinical models reported in the literature. Subsequently, we developed a new model for predicting difficult intubation.

# **Methods**

This protocol was approved by the IRB of the University of Iowa Hospitals and Clinics, Iowa City, IA, and each patient gave his or her written informed consent to participate in the study. Adult patients presenting to the University of Iowa Hospitals and Clinics for general anesthesia for any type of nonemergency surgical procedures except traumatic facial abnormalities, obstetric surgery, or cardiac surgery from October 1999 to November 2004 were enrolled. Unanticipated difficult intubation was identified by an experienced laryngoscopist (>5 yr in anesthetic practice after completion of training). For the purposes of this study, unanticipated difficult intubation was defined as difficult laryngoscopy (corresponding to a Grade 3 or 4 Cormack and Lehane laryngoscopic view) and difficult tracheal intubation (2 or more attempts at placing the endotracheal tube) or the use of an alternative device (laryngeal mask airway [LMA; Laryngeal Mask Company, Henley-on-Thames, United Kingdom] or bougie) when using optimal head and neck positioning (the sniff position). As defined by Cormack and

Lehane (5), Grade 3 means that, during laryngoscopy, none of the glottis, but part or the entire epiglottis, can be seen, whereas Grade 4 means that neither the laryngeal structures nor the epiglottis can be visualized.

Postoperatively, patients who had unanticipated difficult intubation were approached by an investigator after they had fully awoken from general anesthesia. If a patient agreed to participate in the study, an investigator invited a second patient from that day's surgical schedule to participate as a control. Each control patient who had undergone uneventful general anesthesia without any reported difficulties at laryngoscopy or tracheal intubation was closely matched demographically with a study patient to age, weight, height, and sex. Difficult intubation patients and their matched controls were the only patients who consented to participate in this study. A second blinded investigator then evaluated the two patients in the postanesthesia care unit, second-stage recovery facility, or ward. To reduce measurement bias, patients were instructed by the consenting investigator not to comment on their sore throat, potential airway difficulty, or any other aspect of their anesthetic experience to the blinded investigator. The details of the laryngoscopic findings and degree of difficulty of intubation were not known by the investigator who interviewed and measured the patient pairs. All of these assessments were performed by one of three investigators.

The clinical assessment included:

- 1. Measurement of weight, height, and recording of age
- 2. Assessment of the airway according to the pharyngeal structures seen by using the method described by Mallampati et al. (16) with the modification described by Samsoon and Young (6):

Class 1: soft palate, fauces, uvula, and pillars visible

Class 2: soft palate, fauces, and uvula visible Class 3: soft palate and base of uvula visible Class 4: none of the soft palate visible

- 3. Measurement of interincisor gap (in centimeters) with the mouth fully open
- 4. Measurement of thyromental distance (in centimeters) along a straight line from the thyroid notch to the lower border of the mandibular mentum with the head fully extended and mouth closed
- 5. Measurement of thyrosternal distance along a straight line from the thyroid notch to the upper border of the manubrium sterni with the head fully extended and mouth closed
- 6. Measurement of neck circumference
- 7. Determination of the Wilson risk sum score (1), which scores 5 factors (weight, head and neck

**Table 1.** Wilson Risk Sum Score (1)

Risk factor	Level	Variable
Weight	0	<90 kg
0	1	90–110 kg
	2	>110 kg
Head and neck	0	>90°
movement	1	About 90° (i.e., $\pm 10^{\circ}$ )
	2	<90°
Jaw movement	0	IG $\geq 5$ cm or SLux $> 0$
	1	IG < 5  cm and  SLux = 0
	2	IG < 5  cm and $SLux < 0$
Receding mandible	0	Normal
Ū.	1	Moderate
	2	Severe
Buck teeth	0	Normal
	1	Moderate
	2	Severe

IG = Interincisor gap; SLux = Subluxation (maximal forward protrusion of the lower incisors beyond the upper incisors).

movement, jaw movement, receding mandible, and buck teeth) from 0 to 2 for a total range of 0-10 (Table 1)

- 8. Recording history of previous difficulty with laryngoscopy or intubation
- Recording diseases associated with difficulties in laryngoscopy or intubation, such as acromegaly (23) and cervical spondylosis with limitation of neck movements (24).

The sensitivity and specificity of the models described by Wilson et al. (1) (Table 2), Arné et al. (20) (Table 2), and Naguib et al. (21) were assessed based on the data collected. The criteria used to predict difficult tracheal intubation were Wilson risk sum score  $\geq$ 4 or more (1), Arné model score >11 (20), and Naguib model score <0 (21). The Naguib model is based on the formula clinical prediction = 4.9504 + (thyrosternal distance  $\times$  1.1003) + (Mallampati score  $\times$  -2.6076) + (thyromental distance  $\times$  0.9684) + (neck circumference  $\times$  -0.3966).

Positive predictive value and negative predictive value were calculated based on a prevalence of difficult intubation of 5.8% (4), as reported in recent metaanalysis.

Statistical analyses were performed by using the SAS software program (version 9.1; SAS Institute Inc., Cary, NC). Each model was assessed based on the entire group of patients. Thus, the Cochran Q value was computed to test the homogeneity of the patient groups. Demographic differences were determined by using the  $\chi^2$  test and were considered significant when P < 0.05.

We also subjected patient data (age, weight, height, sex, thyromental distance, Mallampati score, interincisor gap, and neck circumference) to a logistic regression model to identify variables that are predictors of difficult intubation. For this analysis, the Mallampati

**Table 2.** Simplified Score Model Described by Arné et al.(20) for Prediction of Difficult Intubation

Risk factor	Score
Previous knowledge of difficult intubation	
No	0
Yes	10
Diseases associated with difficult intubation	
No	0
Yes	5
Clinical symptoms of airway pathology	
No	0
Yes	3
IG and mandible subluxation	
$IG \ge 5 \text{ cm or } SLux > 0$	0
IG < 5.0-3.5 cm and $SLux = 0$	3
IG < 3.5 cm and $SLux < 0$	13
Thyromental distance	
$\geq 6.5 \text{ cm}$	0
< 6.5 cm	4
Maximum range of head and neck movement	
More than 100°	0
About 90° (±10°)	2
Less than 80°	5
Mallampati score	
Class 1	0
Class 2	2
Class 3	6
Class 4	8
Total possible	48

IG = interincisor gap; SLux = subluxation (maximal forward protrusion of the lower incisors beyond the upper incisors).

score was dichotomized such that a score of 1 or 2 was scored as 0 and a score of 3 or 4 was scored as 1.

The receiver operating characteristic (ROC) curve was used to describe the discrimination abilities and to explore the trade-offs between the sensitivity and specificity of the different models (25). The ROC area under the curve (AUC) is frequently viewed as a robust indicator of the performance of classification models. The AUC is a performance indicator equivalent to the nonparametric concordance measure, Somers D, and the difference between two ROC areas is half the difference between the corresponding Somers D values (26). The STATA software program (version 8; Stata Corp, College Station, TX) was used to assess the difference between ROC AUCs based on the  $\chi^2$  test developed from the generalized *U*-statistics theory by DeLong et al. (27).

#### Results

Data were collected on a convenience sample of all patients who presented to the operating rooms at the University of Iowa Hospitals and Clinics between September 1999 and November 2004. During the study period, 210 patients were identified and consented to participate in this study; 194 were included in the final analysis (97 with a difficult airway and 97 controls).

Table 3.	Univariate	Analysis	of Clinical	Variables
I HOIC U.	Omvariate	1 111011 y 010	or chinem	1 al labico

	Difficult laryngoscopy and intubation group (n = 97)	Control group $(n = 97)$	<i>P</i> -value
Age, yr (mean $\pm$ sD)	$54.8 \pm 14.2$	$52.6 \pm 16.4$	0.32
Sex ratio (male/female)	58/39	52/45	0.38
Weight, kg (mean $\pm$ sp)	$88.6 \pm 19.0$	$85.9 \pm 22.2$	0.38
Height, cm (mean $\pm$ sD)	$170.9 \pm 12.2$	$170.4 \pm 12.2$	0.79
IG, cm (mean $\pm$ sD)	$3.6 \pm 0.7$	$4.6 \pm 0.9$	< 0.0001
Thyromental distance, cm (mean $\pm$ sp)	$6.86 \pm 1.3$	$7.95 \pm 1.00$	< 0.0001
Thyrosternal distance, cm (mean $\pm$ sp)	$8.0 \pm 1.5$	$8.7\pm2.0$	0.01
Neck circumference, cm (mean $\pm$ sp)	$42.5 \pm 4.9$	$40.9 \pm 4.7$	0.03
Number of patients per Mallampati score (%)			
Class 1	7 (7.2)	51 (52.6)	
Class 2	25 (25.8)	39 (40.2)	< 0.0001
Class 3	54 (55.7)	4 (4.1)	
Class 4	11 (11.3)	3 (3.1)	

IG = interincisor gap.

Sixteen patients (eight with a difficult airway and eight controls) were excluded because of early discharge, incomplete data, or exclusion criteria. The total patient sample yielded 80% power to detect significant differences in model accuracy (odds ratio, 2.209; 30% model discordance) with the use of a two-sided McNemar test with a significance level of 0.05.

There were no significant differences in the mean age, weight, or height between the two groups (Table 3). However, the mean interincisor gap, thyromental distance, thyrosternal distance, neck circumference, and Mallampati score differed significantly.

In the 97 patients in the difficult intubation group, tracheal intubation was achieved under direct laryngoscopy after several attempts (mean  $\pm$  sp, 3.3  $\pm$  1.1) in 40 patients and with the use of a gum-elastic bougie in another 19 patients. Direct laryngoscopic intubation was completely unsuccessful in 38 of 97 patients. In 15 of these 38 patients in whom direct laryngoscopic intubation was unsuccessful, fiberoptic-guided tracheal intubation was performed successfully (9 while the patient was awake [awakened after intubation failed] and 6 while the patient was asleep, including 1 who underwent fiberoptic-guided tracheal tube placement via an intubating LMA). Fiberoptic-guided intubation was unsuccessful in another five patients. In the remaining 23 of 38 patients, tracheal intubation was performed with the aid of an intubating LMA in 9 patients. A LMA was used in another nine patients and blind nasal intubation in four patients. One patient in whom tracheal intubation was difficult was allowed to awaken, and a regional technique was used. This patient suffered postoperative oral trauma and swelling. No other complications were noted.

The number of patients enrolled in the study during the first 15 months from September 1999 through December 2000 was 86 and decreased thereafter to 32, 18, 30, and 28 patients in December 2001, December 2002, December 2003, and November 2004, respectively.

The highest sensitivity was achieved with the Naguib model (Table 4). Specifically, the sensitivity of this model was 81.4% (95% confidence interval [CI], 74.0%-89.0%) compared with 40.2% (95% CI, 30.0%-50.0%) for the Wilson model and 54.6% (95% CI, 45.0%–65.0%) for the Arné model. Naguib model was significantly more sensitive than the other 2 models based on a pair-wise comparison using the McNemar test (P < 0.0001). Cochran Q statistic value indicated that the 3 models differed significantly with respect to their prediction accuracy (P < 0.02). Both Naguib model and Arné models classified more intubations correctly (P = 0.01) than the Wilson model (Table 5). The McNemar test indicated that the Arné model and Naguib model did not differ significantly in their prediction accuracy (P = 0.6; kappa = 0.2). However, the specificity of the Arné model (94.9% [95% CI, 90.0%-99.0%]) and Wilson model (92.8% [95% CI, 88.0%-98.0%]) was significantly higher (P < 0.0001) than that of the Naguib model (72.2% [95% CI, 63.0%–81.0%).

The ROC AUC that measured the discriminating power of the Arné, Naguib, and Wilson model was 0.87 (95% CI, 0.82–0.92), 0.82 (95% CI, 0.76–0.88), and 0.79 (95% CI, 0.72–0.85), respectively (Fig. 1). The ROC AUC for the Arné model was significantly greater than that of the Wilson model (P = 0.001).

Logistic regression analysis identified four risk factors correlated with the prediction of difficult laryngoscopy and intubation: thyromental distance, interincisor gap, height, and Mallampati score. The prediction (l) was determined by the equation

 $l = 0.2262 - 0.4621 \times$  thyromental distance + 2.5516 × Mallampati score - 1.1461 × interincisor gap + 0.0433 × height,

in which the thyromental distance, interincisor gap, and height were measured in centimeters and Mallampati score was 0 or 1. Using this equation for

Model	Sensitivity (%)	Specificity (%)	Positive predictive value (%)	Negative predictive value (%)
Wilson model	40.2 (95% CI, 30.0-50.0)	92.8 (95% CI, 88.0-98.0)	25.6	96.2
Arné model	54.6 (95% CI, 45.0-65.0)	94.9 (95% CI, 90.0–99.0)	39.7	97.1
Naguib model	81.4 (95% CI, 74.0-89.0)*	72.2 (95% CI, 63.0–81.0)*	15.3	98.4

Table 4. Sensitivity, Specificity, Positive Predictive Value, and Negative Predictive Value of the Three Models Tested

The positive and negative predictive values were calculated based on an overall incidence of difficult intubation of 5.8%, as reported in a recent meta-analysis (4). CI = confidence interval. \*P < 0.0001 as compared with the Wilson and Arné models.

**Table 5.** Prediction Accuracy of the Three Models Tested

Model	Prediction	Number	Percentage
Wilson model	Correct	129	66.5
	Wrong	65	33.5
Arné model	Correct	145	74.7
	Wrong	49	25.3
Naguib model	Correct	149	76.8
Ũ	Wrong	45	23.2

Cochran Q statistic value indicated that the Naguib and Arné models were significantly more accurate (P = 0.01) than Wilson model.



**Figure 1.** Receiver operating characteristic (ROC) curves depicting the relationship between the sensitivity and specificity of the three models of predicting difficult intubation that we tested and our new model. The ROC AUCs show the discriminating power of the models.

predicting difficult intubation, the laryngoscopy and intubation would be easy if the numerical value (*l*) in the equation is less than zero (i.e., negative) but difficult if the numerical value (*l*) is more than zero (i.e., positive).

The posterior probability of group membership for each patient was used to compare the model prediction with the actual outcome. This new model correctly predicted 84% (163 of 194) of the cases. The sensitivity, specificity, positive predictive value, and negative predictive value of this model were 82.5% (95% CI, 73%–89%), 85.6% (95% CI, 77%–91%), 26.1%, and 98.8%, respectively. The ROC AUC for this model was 0.90 (95% CI, 0.86–0.95) (Fig. 1).

A variable correlation analysis showed that height was significantly correlated with both interincisor gap (P < 0.0001) and thyromental distance (P = 0.0007). The existence of this multi-colinearity allows height, which is not significant at the univariate level (Table 5), to be a significant factor in the multivariate model (Table 6).

The total number of adult patients who underwent general anesthesia and were initially eligible for the study during the study period was 73,696. The trachea proved unexpectedly difficult to intubate in 97 patients (0.13%) and was impossible to intubate in 38 patients (0.05%).

## Discussion

This is the first validation study to evaluate different multivariate models in the same population of patients in determining the most sensitive model for prediction of difficult intubation. There was a substantial difference in sensitivity and specificity among the three models tested. The results show that the multivariate model described by Naguib et al. (16) is the most sensitive (P < 0.0001) in identifying patients with unanticipated difficult intubation (81.4% [95% CI, 74.0%–89.0%]). Both the model described by Arné et al. (20) and that described by Wilson et al. (1) had lower sensitivity (54.6% [95% CI, 45.0%-65.0%] and 40.2% [95% CI, 30.0%-50.0%], respectively. However, the specificity of both the Arné model (94.9% [95% CI, 90.0%-99.0%]) and Wilson model (92.8% [95% CI, 88.0%-98.0%) was significantly higher (P < 0.0001) than that of the Naguib model (72.2% [95% CI, 63.0%-81.0%]). Both the Naguib and Arné models were significantly more accurate at correctly identifying easy or difficult intubations (P = 0.01) than the Wilson model was (76.8%, 74.7%, and 66.5%, respectively).

The ideal model for prediction of difficult intubation would have perfect sensitivity and specificity. Sensitivity and specificity are dependent on each other: an increase in one of them usually results in a decrease in the other. High specificity may also increase the positive predictive value despite low sensitivity, as seen with the Wilson and Arné models in this

Table 6. Risk Factors that Correlated with the Predication
of Difficult Intubation as Identified by Multivariate
Analysis in the New Model

Variable	$\chi^2$ (1 DF) <i>P</i> -value	Odds ratio
Thyromental distance	0.0218	0.630
Mallampati score	< 0.0001	12.827
Interincisor gap	0.0005	0.318
Height	0.0118	1.044

Mallampati score was dichotomized such that a Mallampati score of 1 or 2 was scored as 0 and a Mallampati score of 3 or 4 was scored as 1. DF = degrees of freedom.

study. A more pressing question seems to be whether sensitivity and specificity are equally important. Clinical models used to predict difficult tracheal intubation have different trade-offs in optimizing sensitivity and optimizing specificity. We believe that the purpose of any such model should be detection of as many patients with a difficult airway as possible to minimize the potentially serious consequences of unanticipated difficult tracheal intubation. To that end, a model with high sensitivity, rather than high specificity, is required. A model with high sensitivity, low specificity, and low positive predictive value (as seen with the Naguib model) would incorrectly classify patients as having a difficult airway. This would probably increase the financial and emotional costs for the patients when, for example, an alternate intubation technique such as awake fiberoptic-guided intubation, is used. However, these costs may only be a fraction of those that accompany the potentially serious outcome of unanticipated difficult tracheal intubation. Therefore, the sensitivity of a prediction model is more important than the specificity and should be weighted more heavily when determining which model to use.

The value of the simplified risk index used in this study (>11) was the value recommended by Arné et al. (20). In their original description of their model, Naguib et al. (21) reported a sensitivity and specificity of 95% and 91%, respectively, whereas Arné et al. (20) reported a sensitivity and specificity of 94% and 93%, respectively. Of note is that in the present study both models had a lower sensitivity than previously demonstrated.

Oates et al. (17) evaluated the Wilson risk sum (score  $\geq$ 2) in 675 cases. They reported a positive predictive value of 8.9% with a low sensitivity (42%) and high specificity (92%). Using the same threshold, Yamamoto et al. (28) reported that the Wilson risk sum yielded a low positive predictive value (5.9%), low sensitivity (55.4%), and high specificity (86.1%). Similarly, Siddiqi and Kazi (22) reported that both Wilson risk sum (score  $\geq$ 2) and Mallampati classification have a similar sensitivity of 42% but different positive predictive values of 11% and 5%, respectively. A higher threshold is preferred for two reasons. First, the

prevalence of unanticipated difficult tracheal intubation is small. Second, a false-positive result increases the potential for the serious consequence of failed tracheal intubation.

In the present study, univariate differences between the difficult intubation and control groups in the interincisor gap, thyromental distance, thyrosternal distance, neck circumference, and Mallampati score were noted. The most popular clinical test for predicting the ease of tracheal intubation is the Mallampati test (16). Because difficult laryngoscopy is a multifactorial problem, clearly no simple predictive test can be used alone. Simple bedside tests such as the Mallampati test (16,29), thyromental distance measurement (18), and sternomental distance measurement (30,31) have been found to be of limited use in predicting difficult laryngoscopy when performed alone. Effective prediction requires a combination of tests (4,32). A recent meta-analysis found the combination of the Mallampati test and thyromental distance to be the most accurate predictors of difficult intubation; however, this combination has a very low sensitivity of 36% (95% CI, 14%–59%) (4).

We developed a new clinical prediction model that considers the thyromental distance, Mallampati score, interincisor gap, and height. This model is 82.5% sensitive and 85.6% specific with an AUC of 0.90. Height was found to be significantly correlated with both interincisor gap (P < 0.0001) and thyromental distance (P = 0.0007). The significance of height as a predictor of difficult intubation was addressed previously by Schmitt et al. (33). They reported that the ratio of height to thyromental distance was a more sensitive indictor of difficult intubation than the thyromental distance alone (33). We considered a model that included the ratio of height to thyromental distance, which yielded identical results to the new model. For the sake of parsimony, we chose to include only height in our model instead of the ratio of height to thyromental distance, as suggested by Schmitt et al. (33). The new model must be prospectively validated.

The incidence of unanticipated difficult tracheal intubation in this study (0.13%) is less frequent than the range of 1%-18% reported by others (1-3). Also, the incidence of impossible tracheal intubation in our study (0.05%) is at the low end of the range of 0.05%-0.35%reported previously (5,6). The number of patients enrolled in the first 15 months of our study was 44% of the total number of patients (86 of 194). However, the number of patients enrolled decreased dramatically over approximately the next four years, suggesting a possible Hawthorne effect. The authors feel that this study may have increased practitioner awareness of difficult airways in patients presenting for surgery and prompted more aggressive use of alternate airway-management techniques, leading to a decrease in the incidence of unanticipated difficult intubation. A Hawthorne effect (identified observer effect) is defined as the tendency of

individuals to improve their behaviors or performance when they know that they are under observation (34,35). The authors also realize they were dependent upon many individual practitioners' assessment and selfreporting of difficult airways. This is a known limitation of voluntary reporting techniques for critical incidents in health care (36).

A potential limitation of a matched case-controlled study (such as this study) is the possibility that some segments of population may not be adequately represented in the study participants. Another limitation of our study was that it was not a truly prospective study. It can be best described as a quasi-prospective evaluation of three models for the prediction of unanticipated difficult intubation, because patients were identified, recruited, and examined after attempted intubation. However, we do not believe that this would have a significant impact on our results.

In conclusion, our study is the first to provide an evidence-based foundation for selection of the most sensitive model for prediction of unanticipated difficult tracheal intubation. We confirmed the high sensitivity of the Naguib model but failed to do so for the Arné and Wilson models. We also created a new model for predicting unanticipated difficult intubation, although it has not yet been prospectively tested. This model may be more sensitive and specific than any of the models currently used to predict difficult intubation.

## References

- 1. Wilson ME, Spiegelhalter D, Robertson JA, Lesser P. Predicting difficult intubation. Br J Anaesth 1988;61:211–6.
- 2. Benumof JL. Management of the difficult adult airway: with special emphasis on awake tracheal intubation. Anesthesiology 1991;75:1087–110.
- 3. Tse JC, Rimm EB, Hussain A. Predicting difficult endotracheal intubation in surgical patients scheduled for general anesthesia: a prospective blind study. Anesth Analg 1995;81:254–8.
- Shiga T, Wajima Z, Inoue T, Sakamoto A. Predicting difficult intubation in apparently normal patients: a meta-analysis of bedside screening test performance. Anesthesiology 2005;103: 429–37.
- 5. Cormack RS, Lehane J. Difficult tracheal intubation in obstetrics. Anaesthesia 1984;39:1105–11.
- Samsoon GL, Young JR. Difficult tracheal intubation: a retrospective study. Anaesthesia 1987;42:487–90.
- 7. Benumof JL. Difficult laryngoscopy: obtaining the best view. Can J Anaesth 1994;41:361–5.
- 8. Confidential enquiry into maternal and child health: why mothers die 2000–2002—the sixth report of the confidential enquiries into maternal death in the United Kingdom. London: RCOG Press, 2004.
- 9. King TA, Adams AP. Failed tracheal intubation. Br J Anaesth 1990;65:400-14.
- Caplan RA, Posner KL, Ward RJ, Cheney FW. Adverse respiratory events in anesthesia: a closed claims analysis. Anesthesiology 1990;72:828–33.
- 11. Boelle PY, Garnerin P, Sicard JF, et al. Voluntary reporting system in anaesthesia: is there a link between undesirable and critical events? Qual Health Care 2000;9:203–9.

- Johnson KG, Hood DD. Esophageal perforation associated with endotracheal intubation. Anesthesiology 1986;64:281–3.
- Biboulet P, Aubas P, Dubourdieu J, et al. Fatal and non fatal cardiac arrests related to anesthesia. Can J Anaesth 2001;48: 326–32.
- 14. Rose DK, Cohen MM. The airway: problems and predictions in 18,500 patients. Can J Anaesth 1994;41:372–83.
- 15. Practice guidelines for management of the difficult airway: a report by the American Society of Anesthesiologists Task Force on Management of the Difficult Airway. Anesthesiology 1993; 78:597–602.
- 16. Mallampati SR, Gatt SP, Gugino LD, et al. A clinical sign to predict difficult tracheal intubation: a prospective study. Can Anaesth Soc J 1985;32:429–34.
- Oates JD, Macleod AD, Oates PD, et al. Comparison of two methods for predicting difficult intubation. Br J Anaesth 1991; 66:305–9.
- Butler PJ, Dhara SS. Prediction of difficult laryngoscopy: an assessment of the thyromental distance and Mallampati predictive tests. Anaesth Intensive Care 1992;20:139–42.
- 19. el-Ganzouri AR, McCarthy RJ, Tuman KJ, et al. Preoperative airway assessment: predictive value of a multivariate risk index. Anesth Analg 1996;82:1197–204.
- Arné J, Descoins P, Fusciardi J, et al. Preoperative assessment for difficult intubation in general and ENT surgery: predictive value of a clinical multivariate risk index. Br J Anaesth 1998;80: 140–6.
- Naguib M, Malabarey T, AlSatli RA, et al. Predictive models for difficult laryngoscopy and intubation: a clinical, radiologic and three-dimensional computer imaging study. Can J Anaesth 1999;46:748–59.
- Siddiqi R, Kazi WA. Predicting difficult intubation: a comparison between Mallampati classification and Wilson risk-sum. J Coll Physicians Surg Pak 2005;15:253–6.
- 23. Edge WG, Whitwam JG. Chondro-calcinosis and difficult intubation in acromegaly. Anaesthesia 1981;36:677–80.
- Calder I, Calder J, Crockard HA. Difficult direct laryngoscopy in patients with cervical spine disease. Anaesthesia 1995;50:756–63.
- Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology 1982; 143:29–36.
- Somers RH. A new asymmetric measure of association for ordinal variables. Am Sociol Rev 1962;27:799–811.
- DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. Biometrics 1988;44: 837–45.
- Yamamoto K, Tsubokawa T, Shibata K, et al. Predicting difficult intubation with indirect laryngoscopy. Anesthesiology 1997;86: 316–21.
- 29. Cattano D, Panicucci E, Paolicchi A, et al. Risk factors assessment of the difficult airway: an Italian survey of 1956 patients. Anesth Analg 2004;99:1774–9.
- Savva D. Prediction of difficult tracheal intubation. Br J Anaesth 1994;73:149–53.
- Al Ramadhani S, Mohamed LA, Rocke DA, Gouws E. Sternomental distance as the sole predictor of difficult laryngoscopy in obstetric anaesthesia. Br J Anaesth 1996;77:312–6.
- 32. Bainton CR. Difficult intubation: what's the best test? Can J Anaesth 1996;43:541–3.
- Schmitt HJ, Kirmse M, Radespiel-Troger M. Ratio of patient's height to thyromental distance improves prediction of difficult laryngoscopy. Anaesth Intensive Care 2002;30:763–5.
- 34. Parsons HM. What happened at Hawthorne? Science 1974;183: 922–32.
- 35. De Amici D, Klersy C, Ramajoli F, et al. Impact of the Hawthorne effect in a longitudinal clinical study: the case of anesthesia. Control Clin Trials 2000;21:103–14.
- Figueiras A, Tato F, Fontainas J, et al. Physicians' attitudes towards voluntary reporting of adverse drug events. J Eval Clin Pract 2001;7:347–54.