Reports of Investigation

Predicting difficult intubation: a multivariable analysis

Keyvan Karkouti MD FRCPC MSc,*
D. Keith Rose MD FRCPC,†
D. Wigglesworth BSc,†
Marsha M. Cohen MD FRCPC PhD‡

Purpose: To develop a clinically useful and valid model for predicting difficult laryngoscopic tracheal intubation in patients with seemingly normal airways by adhering to the principles of multivariable model development.

Methods: This was an observational study performed at a tertiary-care teaching hospital. Preoperatively, 444 randomly selected patients requiring tracheal intubation for elective surgery were assessed. In addition, 27 patients in whom tracheal intubation was difficult, but were not assessed preoperatively, were assessed postoperatively. One assessor, blinded to the intubation information, collected the predictor variables. A reliable definition for difficult intubation was used and all attempts were made to eliminate sources of bias. Multivariable modeling was performed using logistic regression and the model was validated using the bootstrapping technique.

Results: Of the 461 patients included in the analysis, 38 were classified as difficult to intubate. Multivariable analysis identified three airway tests that were highly significant for predicting difficult tracheal intubation. These were: 1) "mouth opening", 2) "chin protrusion", and 3) "atlanto-occipital extension". Using these tests, a validated, highly reliable and predictive model is produced to determine the probability of difficult intubation for patients. At a selected probability cut-off value, the model is 86.8% sensitive and 96.0% specific.

Conclusion: A simple and accurate multivariable model, consisting of three airway tests, is produced for predicting difficult laryngoscopic tracheal intubation. Additional studies will be required to determine the accuracy and feasibility of this model when applied to a large sample of new patients by multiple anesthesiologists.

Objectif : Élaborer un modèle valide et utilisable en clinique pour prédire des difficultés d'intubation trachéale laryngoscopique chez des patients qui ont apparemment des voies aériennes normales. Utiliser, pour ce faire, les principes d'élaboration d'un modèle multivariable.

Méthode : Il s'est agi d'une observation réalisée dans un hôpital d'enseignement de soins tertiaires. Avant l'opération, on a évalué 444 patients choisis au hasard qui avaient besoin d'intubation endotrachéale pendant une intervention planifiée. De plus, 27 patients chez qui l'intubation a été difficile n'ont été évalués qu'après l'intervention. Un assistant, qui ne connaissait pas les conditions d'intubation, a enregistré les variables de prédiction. Une définition exacte de l'intubation difficile a été utilisée et on a tenté d'éliminer tout biais possible. Une modélisation à multivariables a été réalisée en utilisant une régression logistique et le modèle a été validé par la technique de l'amorce («bootstrapping»).

Résultats : Des 461 patients inclus dans l'analyse, 38 ont été difficiles à intuber. L'analyse à multivariables a reconnu trois épreuves d'intubation comme hautement significatives pour prédire une intubation endotrachéale difficile : 1) «L'ouverture de la bouche», 2) «la protrusion du menton» et 3) «l'extension atlanto-occipitale». Avec ces tests, un modèle validé, très fiable et prédictif a été produit pour déterminer la probabilité d'intubation difficile. Pour une valeur limite de probabilité choisie, le modèle affichait une sensibilité de 86,8 % et une spécificité de 96,0 %.

Conclusion : Un modèle à multivariables simple et précis, fait de trois test d'intubation, a été produit pour prédire des difficultés d'intubation endotrachéale laryngoscopique. D'autres études demeurent nécessaires pour évaluer la fidélité et la faisabilité de ce modèle quand il est appliqué à un échantillon important de nouveaux patients par différents anesthésiologistes.

From the Departments of Anaesthesia, Toronto General Hospital, University Health Network,* and St. Michael's Hospital,† and the Center for Research in Women's Health, Department of Health Administration, and Department of Anaesthesia, Sunnybrook Health Science Centre,‡ University of Toronto, Canada. Work undertaken at St. Michael's Hospital.

Address correspondence to: Dr. K. Karkouti, Toronto General Hospital, University Health Network, Bell Wing 4-645, 200 Elizabeth Street, Toronto, Ontario, M5G 2C4 Canada Email: keyvan.karkouti@uhn.on.ca

Supported by the physicians of Ontario through the Physicians' Services Incorporated Foundation. Dr. Cohen is the recipient of a Senior Scientist Award from MRC Canada.

Presented in part at the Canadian Anaesthetists' Society meeting, 1997, Vancouver, Canada.

Accepted for publication March 14, 2000.

THE failure to maintain a patent airway following the induction of general anesthesia is a major concern for anesthesiologists. For securing the airway, tracheal intubation using direct laryngoscopy remains the method of choice in most cases. However, direct laryngoscopic intubation is difficult in 1% - 4% and impossible in 0.05% - 0.35% of patients who have seemingly normal airways.[1] The unanticipated difficult laryngoscopic intubation places patients at increased risk of complications ranging from sore throat to serious airway trauma.[2,3] Moreover, in some cases the anesthesiologist may not be able to maintain a patent airway, leading to severe complications such as brain damage or death.[4] Since these risks may be avoided by securing the airway prior to the induction of general anesthesia by alternative methods of tracheal intubation (e.g., fibreoptic bronchoscopy), it would be useful to identify these patients preoperatively.

To aid the anesthesiologist in identifying patients whose tracheas are unexpectedly difficult to intubate by direct laryngoscopy, several non-invasive clinical preoperative airway measures have been described that possess significant associations with difficult intubation.[5–18] However, since the ease of laryngoscopic tracheal intubation depends on several airway elements, no single measure of the airway can be expected to predict difficult intubation accurately, and studies have confirmed the low predictive ability of some of these measures.[15,19–22]

To develop more predictive models, several investigators have examined the relationship of multiple airway measures and difficult intubation.[12,23–31] These studies have had conflicting results in that the predictive ability of the resulting models varies widely: most perform poorly whereas others have reported highly accurate models (Table I). This variability may be explained by differences in study design – such as differences in patient populations, airway measures, or definitions for difficult intubation – or, perhaps more importantly, differences in model development.

Methodological standards for developing clinical prediction models have been described by several authors, and failure to adhere to these standards may lead to prediction models that are not valid or clinically useful.[32–35] Prediction models may be invalid for reasons such as inappropriate selection, definition and assessment of independent or dependent variables, or inappropriate statistical modeling strategies.[32–35] Although a detailed discussion of these factors is beyond the scope of this article, a brief outline is presented below in the context of prediction of difficult tracheal intubation.

Independent variables (airway measures): The ease of laryngoscopic tracheal intubation depends on several airway elements: mandibular movement, mandibular space, neck mobility, and oropharyngeal space.[37] For each of these elements, there are several airway measures available, and their reliability and predictive ability varies widely. A model will be appropriate only if it is developed from a pool of reliable and valid airway measures that represent all these elements. Moreover, there should be explicit criteria for the definition, grading, and measurement of the airway measures and they should be collected without knowledge of the patients' outcome.

Dependent variable (difficult intubation): The outcome definition used should be clinically important. Also, to avoid misclassification of patients, the definition needs to be precise and reliable. Finally, to avoid measurement bias, the outcome should be assessed without knowledge of the independent variables.

Statistical issues: To select and weight the independent variables appropriately in a prediction model, multivariable statistical methods are needed. Appropriate execution and interpretation of multivariable analysis requires adherence to statistical principles relating to factors such as model selection, sample size (5-10 events for every predictor variable included in the model), patient selection, variable elimination, and performance evaluations.

Previous studies have not adhered to these principles, and their prediction models may therefore be misleading or invalid. The goal of this study was to develop a clinically useful and valid prediction model for difficult tracheal intubation by adhering to the principles of model development.

Materials and Methods

Patient selection

Following institutional ethics approval, airway assessment was carried out on two groups of patients over

TABLE I Accuracy of models predicting difficult laryngoscopic tracheal intubation*

Study	*Sensitivity (%)*	*Specificity (%)*
Naguib *et al.*[23]	95	91
Arne *et al.*[31]	94	96
Nath *et al.*[30]	96	82
El-Ganzouri *et al.*[29]	60	94
Jacobsen *et al.*[28]	100	86
Tse *et al.*[27]	5-66	65-99
Descoins *et al.*[26]	100	95
Pottecher *et al.*[24]	70	65
Wilson *et al.*[12]	75	88

* Limited to models using bed-side measurements only

six months at St. Michael's Hospital (a tertiary-care teaching hospital). One group consisted of 444 patients who were randomly selected from the operative schedule (since it was not possible to assess all surgical patients), and were assessed preoperatively by one investigator who was not involved in the anesthetic care of the patients.

The second group consisted of patients who were not assessed by the investigator preoperatively, but were noted to have a difficult tracheal intubation (see definition below) at the time of their surgery (n = 27). These patients, who were included to increase the number of difficult intubations in the sample, were assessed following their surgery. To avoid measurement bias, several steps were taken to ensure that the investigator was blinded to the intubation information of these patients: appropriate patients were identified by a research nurse who reviewed all anesthesia records daily; each patient was matched to a 'dummy' patient with similar demographics but easy tracheal intubation; and both the patient with difficult tracheal intubation and the matched 'dummy' patient were referred to the investigator for assessment. The 'dummy' patients were only used to blind the investigator to the intubation information; they were not included in any of the analyses.

Patients were considered eligible for inclusion in the study if they were to receive, or had received, general anesthesia requiring tracheal intubation for elective surgery. The exclusion criteria included an inability to give consent, age less than 18 yr old, pregnancy, unstable cervical spine, gross anatomical abnormalities of head or neck, or any recent surgery involving the head or neck.

Patient assessment

One anesthesiologist, specifically trained to carry out the airway tests used in this study, assessed all patients. The data from patient interviews and examinations were recorded on a standardized form and were not available to the attending anesthesiologists.

Data collected included patient age; sex; and any medical condition that might affect the airway (e.g., snoring, obstructive sleep apnea, diabetes, rheumatoid arthritis), method of induction, or tracheal intubation (e.g., gastric reflux, morbid obesity).

An objective and detailed airway examination was then carried out on each patient. This included documentation of the patient's height, weight, dentition (whether on not patients had full upper dentures), and airway tests that have previously been shown to be capable of predicting difficult tracheal intubation (see Table II for definition of the tests used in this study).[8–14] Only tests which could be easily completed at the bedside were selected. For each test, the most valid and reliable method of examination was used.[36,37] All tests were completed with the patient in the sitting position. The tests were clearly defined (see Table II), and exact measurements were made using an accurate measuring device (C-THRU™ Inch/Metric Protractor Ruler model B-75).

Method of anesthesia

Anesthesiologists at the study hospital, who did not have knowledge of the independent airway assessor's findings, were responsible for intubation and documenting on the anesthesia record the method of intubation (including the type of blade(s) used, method of induction, and whether tracheal pressure was applied), the best view at laryngoscopy, and the number of laryngoscopy attempts. The method of anesthesia was left to the discretion of the attending anesthesiologist.

Outcome definition

The definition of difficult laryngoscopic tracheal intubation was based on the best laryngoscopic view and the number of laryngoscopy attempts, since it has been shown that using both these parameters improves the reliability of identification of difficult laryngoscopic tracheal intubation.[38] The view at laryngoscopy was graded in the following manner: grade 1 if part of the vocal cords was visible, grade 2 if only the arytenoids were visible, grade 3 if only the epiglottis was visible, and grade 4 if the epiglottis was not visible.[12] Tracheal intubation was classified as 'easy' if the number of laryngoscopy attempts plus the grade of laryngoscopy view was less than or equal to four, and as 'difficult' if it was greater than four (see Table III). In addition, patients were classified as 'difficult' if direct laryngoscopy was unsuccessful, or if they underwent awake tracheal intubation based on a history of difficult laryngoscopic tracheal intubation. For the latter group, they were included in the analysis only if their previous anesthesia records included details of the intubation attempts (i.e., the blade(s) used, the number of laryngoscopy attempts, and the best view at laryngoscopy), allowing classification of the intubation as described above.

Statistical analysis

The statistical analyses were carried out using SAS version 6.12 (SAS Institute Inc., Cary, North Carolina). Baseline characteristics of study patients were summarized in terms of frequencies and percentages for categorical variables and median, 25th and 75th percentiles for continuous variables. Logistic regres-

TABLE II Definitions of airway measures*

Independent variables	*Definitions*
Mouth opening	Patient is asked to open his/her mouth as wide as possible. The distance between the upper and lower incisors (or gingiva in edentulous patients) in the mid-line is measured.[17]
Subluxation	Patient is asked to protrude his/her lower incisors as far as possible. Protrusion is ranked as follows: 1) lower incisors anterior to upper incisors (G), 2) lower incisors not anterior to upper incisors (NG).[12]
Thyromental distance	Patient is asked to extend the head as far as possible, keeping the mouth closed. The straight distance from the inside of the mentum to the thyroid notch is measured.[13]
Chin protrusion	Patient is asked to 1) assume the 'sniffing position' by flexing the neck and then tilting the head up as far as possible (the investigator's hand is placed on the lower neck to assure immobility of the lower cervical spine), and 2) opening the mouth as wide as possible. The edge of a clear ruler is then placed on a line from the tip of the upper incisors to the most anterior part of the thyroid cartilage and the perpendicular distance from the edge of the ruler to the genial tubercle in front of this line is measured to the nearest 0.25 cm.[8] This is a slight modification from the original description by Bellhouse *et al.*[8]
Lower cervical spine flexion	Patient is asked to flex the neck maximally. The distance between the genial tubercle and the sternal notch is measured.
Atlanto-occipital extension	Patient is asked to flex the neck completely keeping the mouth slightly open. The tip of a protractor is placed at the external acoustic meatus and is lined up with the occlusal surfaces of the maxillary teeth. The patient then assumes the 'sniffing position' (as above) and the angle transversed by the occlusal surfaces of the maxillary teeth is measured in degrees.[8]
Oropharyngeal view	Patient is asked to extend the neck, open mouth *maximally*, protrude the tongue, and phonate.[36] The visible view is ranked as follows: 1) good visualization of the soft palate, fauces, uvula, and tonsillar pillars; 2) pillars obscured by the base of the tongue, but posterior pharyngeal wall clearly visible below the soft palate; 3) soft palate and base of the uvula visible; 4) soft palate not visible.[10,11]

* All measures were completed while the patient was in the sitting position, an accurate measuring device was used, and measures were rounded off to the nearest 0.5 centimeters unless stated otherwise.

TABLE III Classification scheme for intubation difficulty

		Number of attempts			
		1	2	3	≥4
Best view	1 (partial view of cords)	2	3	4	5
	2 (arytenoids only)	3	4	5	6
	3 (epiglottis only)	4	5	6	7
	4 (no view of epiglottis)	5	6	7	8

Patients classified as "easy intubation" if combined score ≤4 and as "difficult intubation" if >4

sion modeling techniques were used to examine individual and multiple relations between patient predictor variables and the binary outcome of difficult intubation *vs* easy intubation.

The predictor variables were patient sex, age, weight, height, body mass index, dentition, and the airway tests described in Table II. For the continuous predictor variables, since logistic regression modeling assumes that they are linearly related to the log odds of the outcome event, we graphically examined the shape of their relation with the outcome variable. For the categorical variables, cells containing less than five patients were combined only if clinically appropriate.

Height and weight were missing in four male patients who were bed-ridden. Since none of them was obese on visual inspection, they were assigned the average height and weight for men in this study. Patients in whom any other variables were missing were excluded from the analysis.

Bivariate analysis was carried out to assess the association of the predictor variables with the outcome variable, including all two-way interactions. The chi-squared statistic was used for categorical variables and the Wilcoxon rank-sum test for continuous variables.

All significant (P <0.3 – selected because it is a commonly recommended screening criterion for selection of candidate variables to be used in multivariable analysis)[39] baseline predictor variables and interaction terms were used to derive the multivariable model by bi-directional stepwise selection. After each step, the model's performance was assessed by the Akaike Information Criterion (AIC index), which is based on minimizing the deviance of the model with a penalty for each added variable, and the area under the receiver operating characteristic (ROC) curve (c–index in logistic regression).[40] The model building process was stopped when the AIC did not improve appreciably and the area under the ROC curve did not change significantly (P < 0.05) when more variables were added.

TABLE IV Baseline data and association of categorical variables with difficult intubation

Variable	*Category*	*Easy intubation*	*Difficult intubation*	*% Difficult*	χ^2_{1df} *Statistic*	*P*
Sex	Male	153	24	13.6	10.7	<0.001
	Female	270	14	5.0		
Dentition	FU* plates	55	0	0	5.6	<0.05†
	Other	368	38	9.4		
Subluxation	G‡	375	19	4.8	42.0	<0.001
	NG‡	48	19	28.4		
Oropharynx view	1 & 2‡	374	12	3.1	82.7	<0.001
	3 & 4‡	49	26	34.7		

* FU – Full upper

† Fisher's Exact test

‡ See Table II for definitions

TABLE V Baseline data and association of continuous variables with difficult intubation

Variable	*Easy intubations*		*Difficult intubations*			
	Median	*(25th, 75th)*	*Median*	*(25th, 75th)*	*Z Score**	*P†*
Age (yr)	46.5	(34.1, 59.9)	54.7	(39.7, 66.2)	2.6	<0.01
Weight (kg)	71.0	(62, 83)	74.5	(64, 90)	1.3	0.2
Height (cm)	166	(160, 172)	168.5	(161, 175)	1.2	0.2
BMI ($kg \cdot m^{-2}$)	25.5	(22.8, 29.6)	27.2	(23.4, 30.3)	0.8	0.4
Mouth opening (cm)	4.5	(4.3, 5)	3.8	(3.4, 4.0)	7.4	<0.001
Thyromental distance (cm)	6	(6, 7)	5.5	(4.5, 6.0)	5.9	<0.001
Chin protrusion (cm)	3	(2.8, 3.3)	2.3	(2.0, 2.5)	8.1	<0.001
C-spine flexion (cm)	3	(2, 3)	3	(3, 4)	4.6	<0.001
A-O extension (degrees)	35	(35, 35)	26	(25, 30)	9.3	<0.001

* Wilcoxon rank-sum test

† Use of the t test yielded equivalent *P* values

The model's calibration was assessed by the Hosmer-Lemeshow goodness-of-fit chi-square test (which statistically compares the predicted probability with actual probability within population subgroups; the larger the *P* value, the better the fit), and its predictive accuracy was assessed by the area under the ROC curve (an area of 0.5 indicates no predictive discrimination and an area of 1.0 indicates perfect separation of patients with different outcomes).[40] Bootstrapping technique was then used to obtain an estimate of the bias in the predictive accuracy of the model.[41] For each bootstrap sample, patients were drawn randomly, with replacement, from the original population. For each of 2000 bootstrap samples, the model was refitted using the variables selected by the bivariate analyses and then tested on the original sample to estimate the degree to which the predictive accuracy of the model would be expected to deteriorate when applied to an independent sample of patients. The amount of "over-optimism" in the initial predictive analysis was quantified by measuring the decrease in the area under the ROC curve.

The model was used to calculate the predicted probability of difficult intubation for each patient, and the accuracy of the model's predictions were assessed by calculating the sensitivity, specificity, and positive predictive value (assuming a 2% incidence of difficult intubation) of the model at several probability cutoffs. The model's accuracy at different probability cutoffs was used to identify an appropriate cutoff value for classifying patients as easy or difficult intubation – all patients whose predicted probability value lies above the cutoff would be classified as difficult intubation, and all those below the cutoff as easy intubation. Since the clinical application of a logistic regression model requires a calculator (see appendix), the selected cutoff value was used and the variables were simplified (by categorization where appropriate) to develop a simple nomogram for use in everyday clinical practice. The sensitivity and specificity of the nomogram were also assessed.

Results

Of the 471 patients included in the study, 461 were included in the final analysis. Four patients were excluded due to missing variables. Six patients were excluded because their tracheas were intubated by direct fibreoptic intubation, and the required information from their previous intubation attempts was not available.

Tracheal intubation was classified as difficult in 38 patients (27 of whom were assessed following their surgery). Direct laryngoscopic intubation following the induction of general anesthesia was attempted in 32 of 38, but was unsuccessful in 12. Of these, in 11 the trachea was intubated by fibreoptic bronchoscopy and one through a laryngeal mask airway. The remaining six patients did not have direct laryngoscopic intubation; awake fibreoptic bronchoscopy was the initial method of choice for intubation. However, according to previous anesthetic records, direct laryngoscopic tracheal intubation had been attempted on all six, and they would have been classified as difficult intubation according to the classification scheme used in this study.

For all patients in whom intubation was performed after the induction of general anesthesia, intravenous induction followed by muscle relaxation was employed. The #3 MAC blade was used as the initial intubating blade for all but six patients (in all of whom intubation was easy). For patients whose tracheas were difficult to intubate following the induction of general anesthesia, tracheal pressure was applied to improve the laryngoscopic view in every case.

In Table IV, the rate of difficult intubation and the results of the bivariate analysis for preoperative categorical variables are shown. All four categorical variables - sex, dentition, subluxation, and oropharyngeal view - were associated with difficult intubation.

Table V lists both the descriptive statistics for the continuous variables classified by intubation difficulty and the results of the bivariate analysis. Older patients were more likely to experience difficult tracheal intubation. Height, weight, and BMI were not associated with difficult intubation. All airway factors were strongly associated with difficult intubation.

All factors with a P value <0.3 were included in the multivariable analysis. No interaction terms were found to be significant. Three variables - chin protrusion, mouth opening, and A-O extension - were found to be associated with difficult intubation in multivariable analysis (Table VI). The model is well calibrated (Hosmer-Lemeshow goodness-of-fit test chi-square value was 3.018 (6 df, P = 0.81) and is highly discriminative (area under ROC curve = 0.983). The bootstrap results showed that there was minimal bias in the predictive accuracy of the model: the lower 95th percentile bootstrap confidence limit of the ROC area was 0.965.

Table VII lists the sensitivity, specificity, and the positive predictive value for the model at various probability cutoffs. When a predicted probability of 0.2 is used to indicate difficult intubation (see appendix),

TABLE VI Independent clinical predictors of difficult intubation (final multivariable model)*

Variable	*Unit*	*Adjusted O.R.†*	*95% C.I.‡*
Chin protrusion	Continuous (0.25 cm)	3.3	(2.0 – 5.3)
Mouth opening	Continuous (0.5 cm)	5.3	(2.5 – 11.3)
A-O Extension	Continuous (5 degrees)	6.6	(3.2 – 13.6)

Hosmer-Lemeshow goodness-of-fit value = 3.018 with 6 degrees of freedom (P = 0.81)

Area under ROC curve = 0.983

* Indicates the independent contribution of each variable after adjustment for all other factors on the list. This model represents the most predictive combination.

† The adjusted odds ratio measures the change in odds of difficult intubation for each unit change of the variable (i.e., as each measure gets *smaller*, the risk of difficult intubation is *increased*).

‡ Confidence interval

TABLE VII Comparison of accuracy of model at various predicted probabilities of difficult intubation*

Probability of difficult intubation	*Sensitivity (%)*	*Specificity (%)*	*P.P.V.† (%)*	*N.P.V.‡ (%)*
0.1	89.5	93.4	21.7	99.8
0.2	86.8	96.0	30.7	99.7
0.3	78.9	97.9	43.4	99.6
0.4	78.9	99.1	64.1	99.6
0.5	78.9	99.5	76.3	99.6
0.6	78.9	99.5	76.3	99.6
0.7	76.3	99.5	75.5	99.5
0.8	63.2	99.8	86.6	99.2
0.9	44.7	99.8	82.0	98.9

* As calculated by the logistic regression model

† Positive predictive value, 2% incidence of difficult intubation assumed (which is the anticipated rate in the general population)

‡ Negative predictive value, 2% incidence of difficult intubation assumed

TABLE VIII Application of the logistic regression model to predict difficult direct laryngoscopic tracheal intubation

		Atlanto-Occipital Extension (degrees)							
		0	*5*	*10*	*15*	*20*	*25*	*30*	*35*
Mouth Opening (cm)	*2.5*	6.1	5.7	5.3	4.9	4.5	4.1	3.7	3.3
	3.0	5.7	5.3	4.9	4.5	4.1	3.7	3.4	3.0
	3.5	5.4	5.0	4.6	4.2	3.8	3.4	3.0	2.6
	4.0	5.0	4.6	4.2	3.8	3.4	3.0	2.6	2.3
	4.5	4.7	4.3	3.9	3.5	3.1	2.7	2.3	1.9
	5.0	4.3	3.9	3.5	3.1	2.7	2.3	1.9	1.6
	5.5	4.0	3.6	3.2	2.8	2.4	2.0	1.6	1.2

Find the appropriate intersect for the patient according to patient's mouth opening and atlanto-occipital extension. If chin protrusion (cm) is less than or equal to the value at that intersect, then intubation may be difficult.

the model has a sensitivity of 86.8% and specificity of 96.0%. Assuming a 2% incidence of difficult intubation in the general population, the positive predictive value of the model at this cutoff value is 30.7%, and its negative predictive value is 99.7%. Using the logistic model (see appendix) and the cutoff probability of 0.2, a nomogram was developed by categorizing atlanto-occipital extension and mouth opening (Table VIII). When reapplied to the sample, the sensitivity of the nomogram is 80.0% (slightly lower than the logistic model), and its specificity is 99.5% (slightly higher than the model).

Discussion

In this study, we have shown that three airway tests – mouth opening, chin protrusion and atlanto-occipital extension - can predict difficult laryngoscopic tracheal intubation with a high degree of accuracy. Using these tests, which can be completed quickly and easily at the bedside, we have developed and validated (using bootstrapping) a simple and accurate model for predicting difficult laryngoscopic tracheal intubation in patients with no obvious airway abnormalities.

The logistic regression model can either be used to calculate the predicted probability that a patient's trachea will be difficult to intubate (as described in the appendix), or can be converted into a nomogram (Table VIII) to conveniently classify patients into 'easy' or 'difficult' tracheal intubation groups (albeit with a small loss in accuracy). Despite the model's high accuracy, its positive predictive value (the probability of those classified as difficult intubation actually being difficult) is low. This is due to the low incidence of difficult laryngoscopic tracheal intubation in the general population. Nevertheless, if we were to apply the model to a patient population that has a 2% incidence of difficult direct laryngoscopic tracheal intubation, for every one-thousand patients, only about three cases of difficult intubation would be missed, and roughly 39 patients whose tracheas are easy to intubate would be misclassified as being difficult to intubate. Thus, despite the low positive predictive value, it seems that the model is accurate enough to be of practical value in routine clinical practice.

The model obtained in this study has an accuracy that is between those of previous models (Table I). This discrepancy is likely due to differences in study design and statistical analysis. To ensure that the model developed in this study is a reliable and valid measure of the predictive ability of the available bedside tests, we adhered to appropriate model building standards. For example, we included airway tests that were known to have high reliability and validity, and measured all the important airway parameters that affect the ease of laryngoscopic tracheal intubation. In addition, explicit criteria for the definition, grading, and measurement of the airway measures were decided in advance, and one specifically trained assessor was used. Moreover, a reliable and clinically important definition of difficult intubation was selected, and the airway assessment and intubation were carried out independently of each other. Most importantly, appropriate statistical methods were used to develop and validate the model.

However, certain limitations exist that may decrease the model's predictive ability, or its effectiveness, if it is used in routine clinical practice.

First, we did not attempt to standardize the modes of induction and intubation (i.e., patient position, type of blade used, etc.), and we did not analyze the effect of seniority of the anesthesiologists on the intubation. However, review of the data revealed that the anesthetic and intubating techniques were very consistent, and a staff anesthesiologist was involved for every difficult intubation.

Another limitation may be due to the use of a single assessor to perform the airway tests. Since the tests do not have perfect reliability, the model's accuracy may have been lower if multiple assessors were used. In addition, in routine clinical practice, the reliability of the tests may be even lower since they will be performed by anesthesiologists with varying levels of experience and time constraints. No model would be valid if the tests comprising the model are performed in an unreliable manner. Thus, the accuracy of the model in routine clinical practice needs to be assessed.

Third, although there are many medical conditions (e.g., obstructive sleep apnea, rheumatic fever, ankylosing spondylitis) and some common symptoms (e.g., snoring and hypertension) that may be associated with difficult intubation, we did not include these factors in our analyses for two reasons. First, powerful predictors (e.g., rheumatic fever) are not very common, and common predictors (e.g., snoring) are not very powerful.[42] To include rare predictors in the analysis would require a prohibitively large sample size, and the common weak predictors would be eliminated during multivariable analysis. Second, we expect that as long as the patients' medical condition has not had a severe impact on the airway, such as causing an unstable cervical spine, the condition's impact on the difficulty of intubation will be identified by the airway tests used; thus, the model would still be applicable.

Another possible limitation may exist with the method of patient selection; specifically, the inclusion of patients with difficult tracheal intubation who were not assessed preoperatively. The inclusion of this group could bias the results if the assessor was aware of the intubation information at the time of the airway assessment,[32] or if the airway assessment was obtained retrospectively (e.g., from the patient's chart).[33] However, since the assessor was blinded to the intubation information and the airway measures were obtained prospectively, any resulting bias from the inclusion of this group of patients should be minimal.

Another potential source of bias is the method of classification of the difficult intubation group, since it included some patients who did not have direct laryngoscopy performed on them at the time of surgery – they may therefore have been incorrectly placed into the difficult intubation group. This, however, is highly unlikely, since we only included those who had a well-documented history of difficult intubation by direct laryngoscopy in the past, allowing us to apply the same classification scheme that was used for all other study patients.

Finally, although our model is highly accurate, there are still a considerable number of patients who will be misclassified by it, as determined by its false positive (1 - specificity) and false negative (1 - sensitivity) rates. Since the incidence of difficult intubation is low, the false negative rate translates to a small number of difficult intubations being missed by the model, while the false positive rate translates to a larger number of patients being incorrectly classified as difficult intubations. A false negative result may expose patients to increased perioperative risk (decreased effectiveness), and a false positive result may result in needless intubations by an alternate technique (increased cost). Currently, it is not known how accurate intubation prediction models need to be to make their routine clinical use cost-effective. In any case, the cost-effectiveness of any prediction rule will vary depending on the clinical scenario, based on the potential risk of a missed difficult intubation. For example, incorrectly predicting that a 60 kilogram man undergoing an elective surgical procedure in a fasting state will have an easy tracheal intubation is different than making a similar mistake in a pregnant patient undergoing an emergency surgical procedure. Thus, the decision to use the model and apply its results will need to be made by individual anesthesiologists on a case-by-case basis, taking into account the potential risk of misclassification.

In conclusion, by adhering to appropriate model building principles, we have shown that it is possible to predict difficult intubation accurately. We developed a clinical prediction model that includes three airway tests – mouth opening, chin protrusion, and atlanto-occipital extension – that can be carried out at the bedside. The result of the validation analysis suggests that the model should perform well in other similarly defined patient populations, as long as the variables are measured accurately. Future studies are required to determine the validity and cost-effectiveness of this model when used in routine clinical practice.

Acknowledgments

We would like to thank M. Tan, K. Sykora, A. Basinski, and K. Dattilo for their valuable contributions to this study.

References

1 *Benumof JL.* Management of the difficult adult airway. With special emphasis on awake tracheal intubation. Anesthesiology 1991; 75: 1087–110.
2 *Hirsch IA, Reagan JO, Sullivan N.* Complications of direct laryngoscopy. A prospective analysis. Anesthesiology Review 1990; 17: 34–40.
3 *Johnson KG, Hood DD.* Esophageal perforation associated with endotracheal intubation. Anesthesiology 1986; 64: 281–3.
4 *Caplan RA, Posner KL, Ward RJ, Cheney FW.* Adverse respiratory events in anesthesia: a closed claims analysis. Anesthesiology 1990; 72: 828–33.
5 *Calder I, Calder J, Crockard HA.* Difficult direct laryngoscopy in patients with cervical spine disease. Anaesthesia 1995; 50: 756–63.
6 *Block C, Brechner VL.* Unusual problems in airway management: the influence of the temperomandibular joint, the mandible and associated structures on endotracheal intubation. Anesth Analg 1971; 50: 114–23.
7 *Brechner VL.* Unusual problems in the management of airways: I. Flexion-extension mobility of the cervical vertebrae. Anesth Analg 1968; 47: 362–73.

8 *Bellhouse CP, Doré C.* Criteria of estimating likelihood of difficulty of endotracheal intubation with the Macintosh Laryngoscope. Anaesth Intensive Care 1988; 16: 329–37.
9 *King TA, Adams AP.* Failed tracheal intubation. Br J Anaesth 1990; 65: 400–14.
10 *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.
11 *Samsoon GLT, Young JRB.* Difficult tracheal intubation: a retrospective study. Anaesthesia 1987; 42: 487–90.
12 *Wilson ME, Spiegelhalter D, Robertson JA, Lesser P.* Predicting difficult intubation. Br J Anaesth 1988; 61: 211–6.
13 *Frerk CM.* Predicting difficult intubation. Anaesthesia 1991; 46: 1005–8.
14 *Chou H-C, Wu T-L.* Mandibulohyoid distance in difficult laryngoscopy. Br J Anaesth 1993; 71: 335–9.
15 *Savva D.* Prediction of difficult tracheal intubation. Br J Anaesth 1994; 73: 149–53.
16 *White A, Kander PL.* Anatomical factors in difficult direct laryngoscopy. Br J Anaesth 1975; 47: 468–73.
17 *Aiello G, Metcalf I.* Anaesthetic implications of temporomandibular joint disease. Can J Anaesth 1992; 39: 610–6.
18 *Nichol HC, Zuck D.* Difficult laryngoscopy - the "anterior" larynx and the atlanto-occipital gap. Br J Anaesth 1983; 55: 141–4.
19 *Bellhouse CP.* Predicting difficult intubation (Letter). Anaesthesia 1992; 47: 440–1.
20 *Oates JDL, Macleod AD, Oates PD, Pearsall FJ, Howie JC, Murray GD.* Comparison of two methods for predicting difficult intubation. Br J Anaesth 1991; 66: 305–9.
21 *Rose DK, Cohen MM.* The airway: problems and predictions in 18,500 patients. Can J Anaesth 1994; 41: 372–83.
22 *Randell T.* Prediction of difficult intubation. Acta Anaesthesiol Scand 1996; 40: 1016–23.
23 *Naguib M, Malabarey T, AlSatli RA, Al Damegh S, Samarkandi AH.* Predictive models for difficult laryngoscopy and intubation. A clinical, radiologic and three-dimensional computer imaging study. Can J Anesth 1999; 46: 748–59.
24 *Pottecher Th, Velten M, Galani M, Forrler M.* Comparative value of clinical signs of difficult tracheal intubation in women. (French) Ann Fr Anesth Reanim 1991; 10: 430–5.
25 *Rocke DA, Murray WB, Rout CC, Gouws E.* Relative risk analysis of factors associated with difficult intubation in obstetric anesthesia. Anesthesiology 1992; 77: 67–73.
26 *Descoins P, Arné J, Bresard D, Ariès J, Fusciardi J.* Proposal of a new multifactorial score to predict difficult intubation in ENT and stomatological surgery. A preliminary study. (French) Ann Fr Anesth Reanim 1994; 13: 195–200.
27 *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.
28 *Jacobsen J, Jensen E, Waldau T, Poulsen TD.* Preoperative evaluation of intubation conditions in patients scheduled for elective surgery. Acta Anaesthesiol Scand 1996; 40: 421–4.
29 *El-Ganzouri AR, McCarthy RJ, Tuman KJ, Tanck EN, Ivankovich AD.* Preoperative airway assessment: predictive value of a multivariate risk index. Anesth Analg 1996; 82: 1197–204.
30 *Nath G, Sekar M.* Predicting difficult intubation - a comprehensive scoring system. Anaesth Intensive Care 1997; 25: 482–6.
31 *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.
32 *Wasson JH, Sox HC, Neff RK, Goldman L.* Clinical prediction rules. Application and methodological standards. N Engl J Med 1985; 313: 793–9.
33 *Laupacis A, Sekar N, Stiell IG.* Clinical prediction rules. A review and suggested modifications of methodological standards. JAMA 1997; 277: 488–94.
34 *Concato J, Feinstein AR, Holford TR.* The risk of determining risk with multivariable models. Ann Intern Med 1993; 118: 201–10.
35 *Harrell FE Jr, Lee KL, Mark DB.* Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing error. Stat Med 1996; 15: 361–87.
36 *Lewis M, Keramati S, Benumof JL, Berry CC.* What is the best way to determine oropharyngeal classification and mandibular space length to predict difficult laryngoscopy? Anesthesiology 1994; 81: 69–75.
37 *Karkouti K, Rose DK, Ferris LE, Wigglesworth DF, Meisami-Fard T, Lee H.* Inter-observer reliability of ten tests used for predicting difficult tracheal intubation. Can J Anaesth 1996; 43: 554–9.
38 *Karkouti K, Rose DK, Ferris LE.* Tracheal intubation: how should it be documented? Can J Anaesth 1996; 43: A31.
39 *Hosmer DW, Lemeshow S.* Applied Logistic Regression. New York: John Witley & Sons, 1989.
40 *Feinstein AR.* Multiple logistic regression. *In*: Multivariable Analysis: An Introduction. New Haven: Yale University Press, 1996: 297–330.

41 *Efron B, Tibshirani RJ.* Estimates of bias. *In*: Cox DR, Hinkley DV, Reid N, Rubin DB, Silverman BW (Eds.). An Introduction to the Bootstrap. New York: Chapman & Hall, 1993: 124–40.
42 *Diamond GA.* Future imperfect: the limitations of clinical prediction models and the limits of clinical prediction. J Am Coll Cardiol 1989; 14[Suppl A]: 12–22.

APPENDIX

Mathematical application of the model

To calculate the probability of difficult laryngoscopic intubation for each patient that presents for surgery, the values for the three airway measures are obtained and inserted in the following formula:[40]

- Probability of difficult intubation = 1/[1 + exp ((L)],
- where L = 35.9 + [(3.35 ("Mouth opening" (cm)] + [(0.38 x "Atlanto-Occipital Extension" (degrees)] + [(4.77 x "Chin protrusion" (cm)]

Once the predicted probability of difficult intubation is calculated, one can then classify patients as easy or difficult intubations based on a probability cutoff value selected a-priori. An appropriate probability cutoff value can be identified from the information provided in Table VII. At each probability level, the sensitivity, and specificity of the model differs, with the sensitivity decreasing and the specificity increasing as the probability cutoff is increased. We favour a low cutoff probability (i.e., 0.2) since it provides for a higher sensitivity, meaning that very few patients whose tracheas are difficult to intubate would be missed, while maintaining a relatively high specificity, meaning that an acceptable number of patients are incorrectly identified as having difficult tracheal intubation. At the 0.2 probability cutoff value, the model missed five patients with difficult tracheal intubation and incorrectly classified 17 patients as difficult. In routine clinical practice, assuming a 2% incidence of difficult intubation, for every 1000 patients intubated, approximately three patients whose tracheas are difficult to intubate would be missed and another 39 patients would be incorrectly classified as difficult intubation by this model.