

Prediction of Difficult Tracheal Intubation

Time for a Paradigm Change

Olivier Langeron, M.D., Ph.D.,* Philippe Cuvillon, M.D.,† Cristina Ibanez-Esteve, M.D.,‡
François Lenfant, M.D., Ph.D.,§ Bruno Riou, M.D., Ph.D.,|| Yannick Le Manach, M.D., Ph.D.#

ABSTRACT

Background: It has been suggested that predicting difficult tracheal intubation is useless because of the poor predictive capacity of individual signs and scores. The authors tested the hypothesis that an accurate prediction of difficult tracheal intubation using simple clinical signs is possible using a computer-assist model.

Methods: In a cohort of 1,655 patients, the authors analyzed the predictive properties of each of the main signs (Mallampati score, mouth opening, thyromental distance, and body mass index) to predict difficult tracheal intubation. They built the best score possible using a simple logistic model ($SCORE_{Clinic}$) and compared it with the more recently described score in the literature ($SCORE_{Naguib}$). Then they used a boosted tree analysis to build the best score possible using computer-assisted calculation ($SCORE_{Computer}$).

Results: Difficult tracheal intubation occurred in 101 patients (6.1%). The predictive properties of each sign remain low (maximum area under the receiver operating characteristic curve 0.70). Using receiver operating characteristic curve, the global prediction of the $SCORE_{Clinic}$

What We Already Know about This Topic

- Predicting difficult endotracheal intubation using a combination of simple physical characteristics is poor and, some have argued, nearly useless

What This Article Tells Us That Is New

- In a cohort of 1,665 patients, 6.1% of whom had difficult endotracheal intubation, application of a computer-assisted model using complex interaction yielded excellent prediction
- Whether the model generated from this group of patients can be generally applied to yield similarly excellent prediction awaits further research

(0.74, 95% CI: 0.72–0.76) was greater than that of the $SCORE_{Naguib}$ (0.66, 95% CI: 0.60–0.72, $P < 0.001$) but significantly lower than that of the $SCORE_{Computer}$ (0.86, 95% CI: 0.84–0.91, $P < 0.001$). The proportion of patients in the inconclusive zone was 71% using $SCORE_{Naguib}$, 56% using $SCORE_{Clinic}$, and only 32% using $SCORE_{Computer}$ (all $P < 0.001$).

Conclusion: Computer-assisted models using complex interaction between variables enable an accurate prediction of difficult tracheal intubation with a low proportion of patients in the inconclusive zone. An external validation of the model is now required.

DIFFICULT tracheal intubation (DTI) remains an important cause of anesthesia-related hypoxic brain damage and death although a significant decrease has recently been observed,¹ related to improved airway management through adoption of guidelines with airway management algorithms,^{2,3} widespread use of devices such as gum elastic bougie,⁴ laryngeal mask, fiberoptic intubation, new optical devices,⁵ and recommendation to prioritize oxygenation.⁶ However, among the strategies proposed to decrease morbidity and mortality related to DTI the role of its prediction remains a matter of debate. Although many algorithms include preoperative assessment, some authors have suggested that attempting to predict DTI is unlikely to be useful.⁷

Several clinical signs have been identified as predictors of difficult laryngoscopy or DTI, including the Mallampati score, mouth opening (MO), and the thyromental distance (TMD), and body mass index (BMI).⁸ However,

*Professor, #Assistant Professor, Department of Anesthesiology and Critical Care, ||Professor and Chairman, Department of Emergency Medicine and Surgery, Université Pierre et Marie Curie-Paris, CHU Pitié-Salpêtrière, Paris, France. †Staff Anesthesiologist, Department of Anesthesiology and Pain Management, CHU Carémeaux, Nîmes, France. ‡Staff Anesthesiologist, Department of Anesthesiology and Critical Care, CHU Pitié-Salpêtrière, Paris, France. §Staff Anesthesiologist, Department of Anesthesiology and Critical Care, CH de Beaune, Beaune, France.

Received from the Department of Anesthesiology and Critical Care, and the Department of Emergency Medicine and Surgery, Université Pierre et Marie-Curie-Paris 6, Centre Hospitalo-Universitaire (CHU) Pitié-Salpêtrière, Assistance Publique-Hôpitaux de Paris, Paris, France; Department of Anesthesiology and Pain Management, CHU Carémeaux, Nîmes, France; and Department of Anesthesiology and Critical Care, Centre Hospitalier (CH) de Beaune, Beaune, France. Submitted for publication June 28, 2012. Accepted for publication September 7, 2012. Support was provided solely from institutional and/or departmental sources.

Address correspondence to Dr. Langeron: Department of Anesthesiology and Critical Care, CHU Pitié-Salpêtrière, 47–83 Boulevard de l'Hôpital, 75651 Paris cedex 13, France. olivier.langeron@psl.aphp.fr. Information on purchasing reprints may be found at www.anesthesiology.org or on the masthead page at the beginning of this issue. ANESTHESIOLOGY's articles are made freely accessible to all readers, for personal use only, 6 months from the cover date of the issue.

Copyright © 2012, the American Society of Anesthesiologists, Inc. Lippincott Williams & Wilkins. Anesthesiology 2012;117:1223-33

the sensitivity and predictive positive values of these signs are low, precluding an accurate prediction of DTI and this has been confirmed by a recent meta-analysis.⁹ Several studies have been proposed to derive a score from multivariate analysis. Although the predictive properties of these scores were higher than those of individual signs, they remain not very high.^{10–12} Moreover, several methodological weaknesses were present in these previous studies: (1) some scores were developed to predict difficult laryngoscopy and not DTI;¹⁰ (2) absence of assessment of the calibration of the score; (3) absence of internal or external validation;¹¹ and (4) use of a dichotomous approach although a separation into three classes (high, intermediate, and low risk) is now recommended.¹³ Most of these studies did not verify that a logistic model accurately depicts the relationship between variables and outcomes. Last, the possibility that interaction between variables may play an important role was ignored although the key issue might be the proportionality of variables and not the variables themselves: a small individual (with a small MO and TMD) may not be difficult to intubate. In contrast to some authors,^{7,9} we considered that a new paradigm is required for the accurate prediction of DTI, incorporating the more recent developments in biostatistical analysis,¹³ which would provide a valuable tool from a dynamic approach including interactions between variables.

Therefore, we conducted an analysis of variables associated with DTI in a large multicenter cohort. First, we analyzed the predictive properties of each main signs to predict DTI. Second, we tried to build the best score possible using a simple logistic model and compared it with the best score previously described in the literature. Third, we used a more sophisticated technique using computer-assisted calculation. In fact, we consider that such scores could be also determined at the bedside, taking into account the new technologies available. In this study, we test the hypothesis that a better prediction (*i.e.*, higher area under the receiver operating characteristic [ROC] curve) of DTI using simple clinical signs is possible, pending a changing paradigm in this prediction and risk assessment of difficult airway management, taking into account both interactions between variables and the concept of the inconclusive zone (*i.e.*, less patients in the “gray zone”).¹³

Materials and Methods

Institutional review board (Comité de Protection des Personnes Pitié-Salpêtrière, Paris, France) approval was obtained. Because there was no randomization and only routine care was performed, waived informed consent was authorized.

Consecutive adult (aged 18 yr and above) patients scheduled for surgery requiring general anesthesia and tracheal intubation using a standard technique with a laryngoscope were included prospectively in our study, including emergency cases and those requiring rapid sequence induction. Patients undergoing regional anesthesia, those undergoing general anesthesia

without tracheal intubation (face mask ventilation, laryngeal mask) were excluded. We also excluded patients with indication of intubation with fiberoptic intubation or those in whom another method was used first (videolaryngoscope, fiberoptic intubation). All patients were included over a 19-week period (February 13, 2012–June 24, 2012) in three centers (Paris, Nîmes, and Beaune, all in France).

Information was collected by the anesthesiologists on a standard form. The following information was collected: age, sex, weight, height, BMI, Mallampati classification as modified by Samssoon and Young¹⁴ performed with the patient in the sitting position with the head in extension, mouth fully opened, tongue out, and with phonation,¹⁵ whenever possible, the TMD measured with the patient in sitting position and head extension,¹⁵ MO measured as the interincisor distance,¹⁶ presence of receding mandible, macroglossia, beard, and lack of teeth. Patients were also asked whether they were habitual (almost every night or every night) snorers or not,¹⁶ without asking for the snoring loudness.

DTI was defined as a proper insertion of the endotracheal tube with conventional laryngoscopy requiring more than two attempts or lasting more than 10 min, or requiring an alternate technique (bougie, videolaryngoscope, Fastrach® [Laryngeal Mask Company Ltd., Le Rocher, Seychelles], fiberoptic).¹⁶ Data collected concerning tracheal intubation were: use of paralyzing agents, scheduled *versus* urgent surgery, Sellick maneuver, characterization of tracheal intubation (easy, difficult, impossible), and grading of the best laryngoscopic view according to the Cormack and Lehane classification.¹⁷ To minimize uncertainty and inaccuracy of the numerical grading system, schematic diagrams were provided for classification of the view of oropharynx and the glottis, according to Mallampati as modified by Samssoon and Young, and to Cormack and Lehane classifications directly in the data chart, as well as a tape-measure (in mm), as previously described.¹⁶ We also recorded the characteristics of the individual who attempt first to intubate (nurse anesthetist, resident in anesthesiology, anesthesiologist).

In our institutions, the routine procedure for tracheal intubation has been standardized. The patient's head and neck were placed in an optimal position (the sniff position)¹⁸ to improve laryngoscopy and intubation outcome. Preoxygenation of each patient during 4 min by bag and mask with 100% oxygen without ventilation was required, and oxygenation was continued during mask ventilation except for rapid sequence in emergency cases. Each patient was routinely monitored during the whole procedure by electrocardiography, pulse oxymetry, and end-tidal carbon dioxide concentration. Because the type of laryngoscopic blade may influence difficulties in tracheal intubation, at least in emergency conditions,^{19,20} all patients were intubated using a standard Macintosh metallic reusable blades (Heine classic+; Heine, Herrshing, Germany). In patients requiring emergency surgery and with a full stomach, a rapid sequence using Sellick maneuver and succinylcholine

(1 mg/kg) was performed.^{19,20} After tracheal intubation, the correct positioning of the endotracheal tube was confirmed by the anesthesiologist using first detection and curve analysis of end-tidal carbon dioxide and then using bilateral auscultation of lungs. When facing a difficult airway, the French Society of Anesthesiology and Critical Care algorithm was followed.³

Diagnosis Properties of Clinical Tests or Scores

We determined the diagnostic properties of each main variable previously described in the literature (MO, TMD, Mallampati class, BMI) using classical diagnosis variables (sensitivity, specificity, positive and negative predictive values) as well as with the positive and negative diagnosis likelihood ratio. The discriminative power of different predictors of difficult intubation was assessed in the study population, using ROC curves. These curves were obtained by averaging 1,000 populations bootstrapped (sampling with replacement) from the original study population. This method limits the impact of outliers and allows the provision of more robust representations. (CI of the average ROC curves were depicted using box plots. Finally, to provide comparison between studies, the (sensitivity, 1- specificity) values calculated in the previous reports were plotted on the ROC curves related to the study population.

Threshold and Inconclusive Zone Determination

A gray zone approach was used to describe the value for which the variables of interest did not provide conclusive information, that is, the inconclusive zone.¹³ This corresponds to a range of values, for which formal conclusions could not be obtained. To determine this interval of values, we first assessed the optimal threshold using Youden index ($J = \text{sensitivity} + \text{specificity} - 1 = \text{true-positive rate} - \text{false-positive rate}$).²¹ Maximizing J corresponds to maximizing the overall correct classification rates and minimizes misclassification rates.¹³ Youden index determination was then conducted for each bootstrapped population, conducting to a set of 1,000 “optimal” values. The mean value of these optimal values and its 95% CI were then estimated. The inconclusive zone was thus defined for each predictor as well as its 95% CI. This first approach was completed by an alternative approach, aiming at defining three classes of response: negative, inconclusive, and positive. We defined inconclusive responses for values presenting with either sensitivity lower than 90% or specificity lower than 90% (10% of diagnosis tolerance). To better present these results, a two-curve (sensitivity and specificity) representation was provided. These two approaches were not exclusive but complementary, and finally, the largest CI was retained as the inconclusive zone.

Construction of the Predictive Models

We used three scores in our study. The first one was the score described by Naguib *et al.*¹² ($\text{SCORE}_{\text{Naguib}}$), which

was considered as the best score previously described in the literature:

$$l = 0.2262 - 0.4621 \cdot \text{TMD} + 2.5516 \cdot \text{Mallampati score} - 1.1461 \cdot \text{MO} + 0.0433 \cdot \text{height}$$

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 greater than zero (*i.e.*, positive).¹² Some previous studies have provided better predictive scores but they included as a risk factor the fact that the patients already experienced difficult intubation,¹¹ and we are of the opinion that these patients should be excluded from any prediction attempt becoming useless by definition, and just be considered as difficult to intubate. The second score was the best score, which could be built from our population but being also simple to calculate at the bedside ($\text{SCORE}_{\text{Clinic}}$), as previously described.²² Last, without consideration for the easiness to construct this prediction score of a difficult airway, we determined the best model feasible using generalized boosting modeling ($\text{SCORE}_{\text{Computer}}$).²³ This approach allows us to take into consideration not only single variables (age, sex, BMI, Mallampati classification, MO, TMD, macroglossia, receding mandible, snoring, were finally retained in the model, see appendix) but also the interactions between variables (as depth as the fourth level) and does not require hypothesis regarding the linearity of the relationship between continuous variables and the endpoint. Generalized boosting modeling is considered outperforming standard logistic regression.²⁴ We used this approach to demonstrate that interactions and nonlinearity of the relationship between predictors and the endpoint are of major importance to predict DTI.

Three steps were followed in the $\text{SCORE}_{\text{Clinic}}$ construction. First, a multiple forward stepwise logistic regression was performed to assess variables associated with difficult tracheal intubation. We used a semiparsimonious approach, and only unbiased variables, which were available (table 1), were included. Discrimination of the final models was assessed by c -statistic and calibration by the Hosmer–Lemeshow statistic. An internal validation was performed using 10-fold crossvalidation.²⁵ Two terms were defined to present the internal validation of the models: the difference (optimism) between the c -statistic observed in the complete population and in the crossvalidation samples; and the optimism corrected c -statistic, expressed as the difference of c -index. Second, we tried to transform the continuous variables selected by the model. Third, we tried to simplify as much as possible the weight allocated to each variable retained in the model. Because ROC analyses have shown an advantage for a score with variable weights compared with a score with equal weights for all variables (data not shown), the weight of each variable included in the score was derived from the logistic

Table 1. Comparison of Patients with and without DTI

	Patients without DTI	Patients with DTI	P Value
	(n = 1,554)	(n = 101)	
Men	793 (51)	62 (61)	0.051
Women	761 (49)	39 (39)	
Age, yr	56 ± 17	58 ± 14	0.15
Age ≥ 75 yr	223 (14)	8 (8)	0.07
Height, cm	169 ± 10	170 ± 10	0.28
Weight, kg	75 ± 18	79 ± 19	0.02
BMI, kg/m ²	26.3 ± 5.6	27.5 ± 6.1	0.049
BMI ≥ 30, kg/m ²	310 (20)	29 (29)	0.04
MO, mm	43 ± 10	38 ± 9	< 0.001
TDM, mm	75 ± 18	71 ± 18	0.02
Mallampati class			
1	873 (56)	25 (25)	<0.001
2	519 (33)	41 (41)	
3	138 (9)	25 (25)	
4	24 (1)	10 (10)	
Receding mandible	262 (17)	35 (35)	<0.001
History of snoring	145 (9)	18 (18)	0.009
Macroglossia	62 (4)	14 (14)	<0.001
Lack of teeth	186 (12)	5 (5)	0.035
Beard	45 (3)	4 (4)	0.54
Sellick maneuver	116 (7)	11 (11)	0.24
Muscle relaxant use	1,442 (93)	87 (86)	0.03
Type of surgery			
Scheduled	1,271 (82)	82 (82)	0.50
Urgent	214 (14)	17 (17)	
Unknown	58 (4)	2 (2)	
Orthopedic	317 (20)	21 (21)	0.27
Abdominal	257 (16)	8 (8)	
Urological	147 (9)	12 (12)	
Vascular and cardiac	129 (8)	8 (8)	
Neurosurgery	384 (25)	27 (27)	
Others	178 (11)	17 (17)	
Unknown	142 (9)	8 (8)	
Cormack and Lehane class			
1	1,211 (78)	2 (2)	<0.001
2	232 (15)	9 (9)	
3	44 (3)	57 (56)	
4	8 (1)	33 (33)	
Unknown	59 (4)	0 (0)	
Number of laryngoscopy*			
1	1,421 (91)	3 (3)	—
2	133 (8)	46 (45)	
> 2	0 (0)	52 (51)	
Number of individuals who attempt to intubate*			
1	1,492 (96)	35 (35)	—
2	62 (4)	55 (54)	
> 2	0 (0)	11 (11)	
First individual who intubates			
Nurse anesthetic	1,158 (75)	63 (62)	0.005
Resident	178 (11)	13 (13)	
Physician	209 (13)	25 (25)	

Data are mean ± SD or number (percent).

* Not tested for significance between patients with and without DTI.

BMI = body mass index; DTI = difficult tracheal intubation; MO = mouth opening; TMD = thyromental distance.

regression coefficients. However, several methods were tested: direct sum of the odds ratio, sum of the logistic coefficient, and the process of converting the logistic regression output into a risk index.²⁶

Because it is important to assess a risk score beyond a dichotomous approach,¹³ the population was divided into three categories: patients at low, intermediate, and high risk for difficult tracheal intubation, the intermediate risk corresponding to the inconclusive zone. To provide an unbiased estimate (and its 95% CI) of the endpoint in each stratum of risk, the bootstrap method was used (random selection with replacement).

Statistical Analysis

The number of patients to be included was calculated to obtain at least 100 events (*i.e.*, 100 DTI). Data are expressed as mean \pm SD and median (95% CI) for non-normally distributed variables, and numbers (percentages) for proportion. Normality was assessed with the D'Agostino–Pearson omnibus test. Comparison of two means was performed using the unpaired Student *t* test, comparison of two medians was performed using the Mann–Whitney test, and comparison of proportions was performed using Fisher exact method. The comparison of two areas under the ROC curves was performed using the nonparametric technique described by DeLong *et al.*²⁷ All *P* values were two-tailed and a *P* value of less than 0.05 was considered significant. Statistical analysis was performed with R software** and specific package.²⁸

Results

Analysis of Main Variables

We included 2,048 consecutive patients. Some patients were excluded because a plastic blade was used ($n = 146$), an alternate technique (videolaryngoscope, fiberoptic) was used at the first attempt ($n = 5$), or age was below 18 yr ($n = 2$). Important missing values occurred in 240 patients. Thus, 1,655 patients were retained in the final analysis (fig. 1). DTI occurred in 101 (6.1%, 95% CI: 5.0–7.4%). Among these patients, DTI was determined because of prolonged intubation (more than 10 min) in 20 patients (20%), more than two attempts in 51 patients (50%), and use of an alternative methods in 74 patients (73%) (the sum differs from 100% because several criteria might occur in the same patients). The alternative method were the use of gum elastic bougie ($n = 46$), videolaryngoscope ($n = 24$), fiberoptic bronchoscope ($n = 6$), and Fastrach® ($n = 2$) (the sum differs from 59 because several techniques might be used in the same patients). No tracheotomy was performed and only one case of impossible tracheal intubation occurred. In this later patient, general anesthesia was performed with laryngeal mask.

Table 1 shows the main characteristics of patients with and without DTI. We determined the diagnostic properties as well as the proportion of patients in and out of the inconclusive zone for each main variable (table 2) previously described in the literature (MO, TMD, Mallampati class, BMI). Figure 2 emphasizes the poor global diagnostic properties of each of them reflected by the low, although significant, area under the ROC curve values and shows the importance of the inconclusive zone.

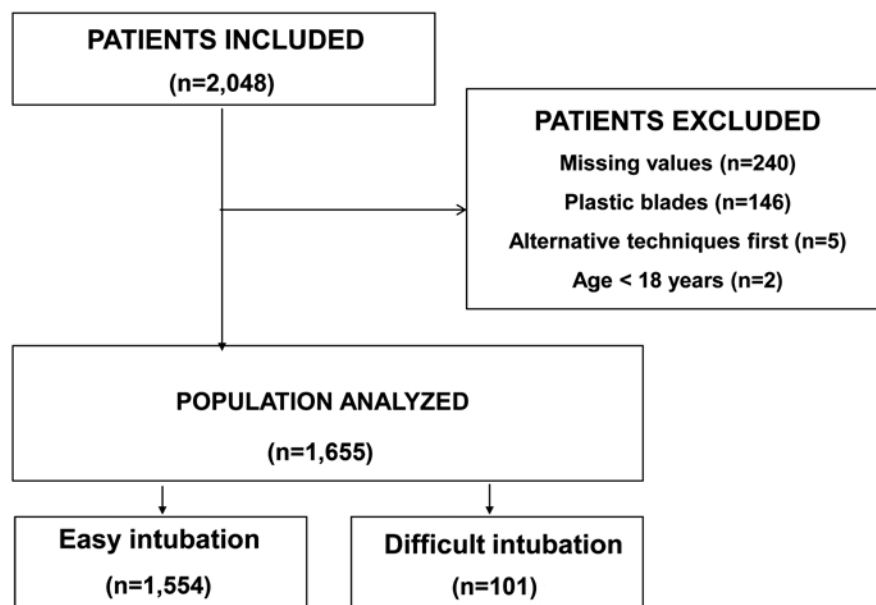


Fig. 1. Flow chart of the study.

** The R Project Contributors. The R project for statistical computing. www.r-project.org. Accessed October 31, 2012.

Table 2. Proportion of Patients in the Inconclusive Zone

	Inconclusive Zone		Patients in the Inconclusive Zone	Patients with Difficult Intubation in the Inconclusive Zone
	Lower Limit	Upper Limit		
MO, mm	25	55	1,220 (90%)	121 (93%)
TMD, mm	50	100	1,224 (90%)	114 (88%)
Mallampati class	1	3	1,327 (98%)	117 (90%)
BMI, kg/m ²	20	34	1,077 (79%)	95 (73%)

BMI = body mass index; MO = mouth opening; TMD = thyromental distance.

Construction and Comparison of Scores

Table 3 reports the model of the logistic regression with continuous variables. Seven independent predictors of difficult tracheal intubation were selected: BMI, Mallampati class, MO, TDM, sex, receding mandible, and the Sellick maneuver. The observed difference between the c-statistic in the entire derivation population and that obtained after 10-fold crossvalidation was below 0.02. Table 3 also reports the results of the logistic regression using stratified variables. The BMI was segmented into three categories (less than 25, 25–34, and more than 35 kg/m²), as well as MO (35 or lesser, 36–50, 50 or more mm), and TDM (60 or lesser, 61–99, 90 or more mm). Mallampati class was entered without any change in the model. The SCORE_{Clinic} was defined using the results of this logistic regression (table 3). This transformation of the variables was associated with a nonsignificant reduction of the c-statistic between the model with continuous variables and the SCORE_{Clinic} (respectively 0.75 [95%

CI: 0.72–0.77] and 0.74 [95% CI: 0.72–0.76], *P* = 0.22) and justified by the ease of obtaining the SCORE_{Clinic} compared with the model with continuous variables.

The global predictive ability of the SCORE_{Clinic} using ROC curve analysis (0.74, 95% CI: 0.72–0.76) was significantly greater than that of the SCORE_{Naguib} (0.66, 95% CI: 0.60–0.72, *P* < 0.001) but significantly lower than that of the SCORE_{Computer} (0.86, 95% CI: 0.84–0.91, *P* < 0.001) (fig. 3). Figure 4 shows the relative influence of each variable (*i.e.*, the reduction of squared error attributable to each variable) used in SCORE_{Computer}. Tracheal intubation was performed without muscular relaxation in 7.6% of our patients, a condition that might favor DTI. Muscle relaxant use was less frequently associated with DTI (5.7% *vs.* 11.1%, *P* = 0.01), but this variable was not significant in the multivariate model. However, we performed a sensitivity analysis excluding these patients, and the results were not significantly modified (data not shown).

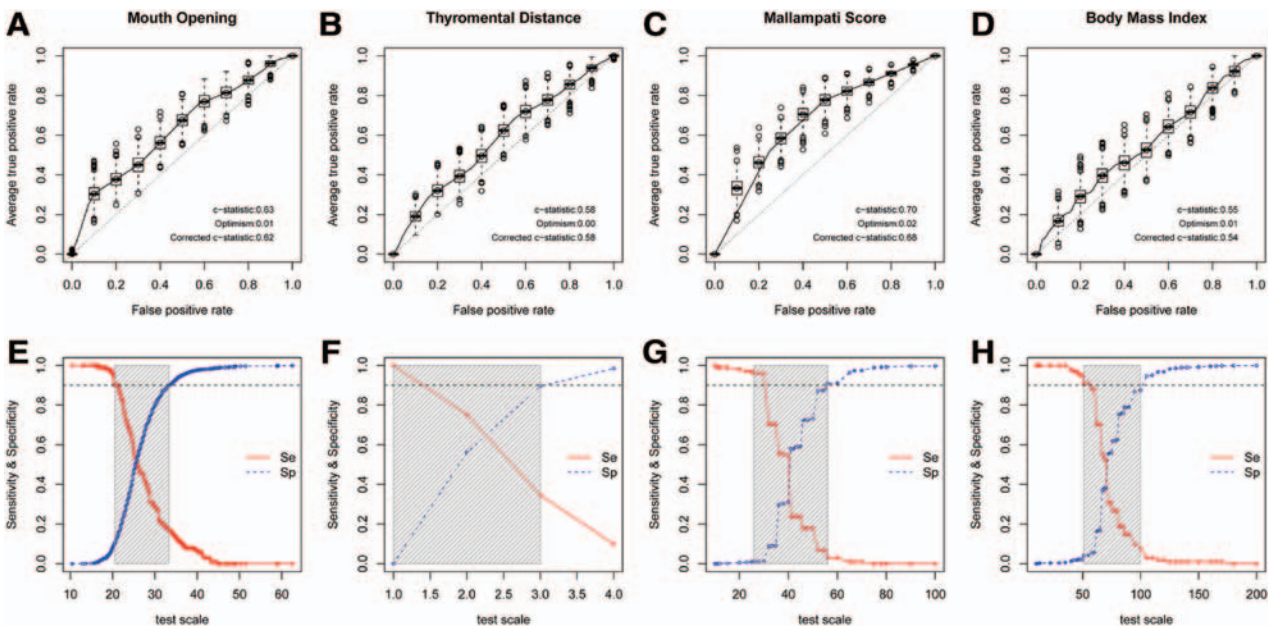


Fig. 2. Average receiver operating characteristic curves obtained by 1,000-fold resampling of the original study population for mouth opening (A), thyromental distance (B), Mallampati’s classes (C), and body mass index (D) (*n* = 1,655). Distribution of the resampled population is represented by box plots. Two graphs ROC curves with determination of the inconclusive zone (*dashed* area) using a diagnosis tolerance of 10% for mouth opening (E), thyromental distance (F), Mallampati’s classes (G), and body mass index (H). Se = sensitivity; Sp = specificity.

Table 3. Multivariate Analysis of Predictors of Difficult Tracheal Intubations and Construction of the SCORE_{Clinic} (n = 1,655)

Variables	Predictors Analyzed as Continuous Variables		Predictors Analyzed as Discontinuous Variables		Points of the SCORE _{Clinic}
	Odds Ratio (95% CI)	P Value	Intervals	Odds Ratio (95% CI)	
BMI, kg/m ²	1.03 (1.00–1.06)	0.05	< 25	1	0
			25–34	0.95 (0.65–1.49)	0
			≥ 35	1.60 (0.76–3.15)	+1
Mallampati class	2.21 (1.74–2.82)	< 0.001	1	1	0
			2	2.58 (1.55–4.38)	+2
			3	4.96 (2.67–9.78)	+4
			4	7.32 (2.84–18.02)	+6
MO, mm	0.97 (0.94–0.99)	0.05	> 50	1	0
			36–50	1.39 (0.66–3.50)	+1
			≤ 35	3.65 (1.59–9.71)	+2
TMD, mm	0.98 (0.97–1.00)	0.05	≥ 90	1	0
			61–89	1.24 (0.63–2.66)	+1
			≤ 60	2.04 (1.30–3.34)	+2
Sex	1	0.04	Woman	1	0
			Man	1.53 (1.01–2.38)	+1
Receding mandible	1	< 0.001	No	1	0
			Yes	2.37 (1.49–3.72)	+2

A low score is associated with a low probability of difficult tracheal intubation (Hosmer–Lemeshow statistic: $\chi^2 = 0.37$; $df = 8$; $P = 0.71$; c-statistic = 0.75).

BMI = body mass index; MO = mouth opening; TDM = thyromental distance.

Determination of the Inconclusive Zone for the Scores

Using the inconclusive zone, we defined three groups of patients: low, intermediate, and high risk for difficult tracheal intubation (fig. 5). Using this approach, 56% of the patients presented values of the SCORE_{Clinic} in the inconclusive zone, whereas 71% of them had inconclusive values of the SCORE_{Naguib} ($P < 0.001$) and 32%, using the SCORE_{Computer} ($P < 0.001$).

Discussion

The goal of our study was to understand better the precise mechanisms involved in the prediction of DTI. We observed that the main variables usually used (MO, TMD, Mallampati class, BMI) are poor individual predictors with a very large inconclusive zone. The diagnostic accuracy of simple scores built using multivariate logistic model is good but probably not sufficient for clinical practice and is associated with a relatively important inconclusive zone. Only a sophisticated model integrating interaction term and taking into consideration the lack of linearity of the main predictors provided an acceptable prediction tool. Boosted trees are the approach we retained in this study, but other similar approaches might also be adapted to predict difficult intubation.

In our study, MO, TMD, and BMI were associated with DTI, as previously reported.^{2,3,8–12} We did not observe that emergency surgery or Sellick maneuver were associated with DTI (table 1). Intubation in emergency conditions

has been reported to be more difficult,^{5,29} although some studies reported opposite results.³⁰ Although cricoid pressure during the Sellick maneuver may increase difficulties in performing laryngoscopy and thus tracheal intubation, it is also the only risk factor that can be immediately modified when facing difficulties by simply releasing it. In our study, the Sellick maneuver was not associated with DTI and our results concord with previous study.³⁰ Most of the patients in our cohort were intubated by a nurse anesthetist, this procedure corresponds to the routine practice in our institutions. However, in the univariate analysis, intubation by a nurse anesthetist or resident was not associated with more frequent DTI (table 1). A low proportion of patients (7.6%) were intubated without muscular relaxation, using local anesthesia of the vocal cords and this practice was significantly associated with DTI (table 1). However, a sensitivity analysis showed that this did not influence our results.

The poor predictive value of each individual variable has been well recognized by previous studies and meta-analysis.^{7,9} Our study confirms this issue and helps us to understand why this occurs. Importantly, the inconclusive zone was wide for each of these variables and the relationship between these variables and the outcome could not be simply described by a linear relationship (fig. 2). Several previous studies demonstrated that a multivariate composite risk score performs better than individual variables do.^{10–12,31} However, these previous studies used simple logistic models,

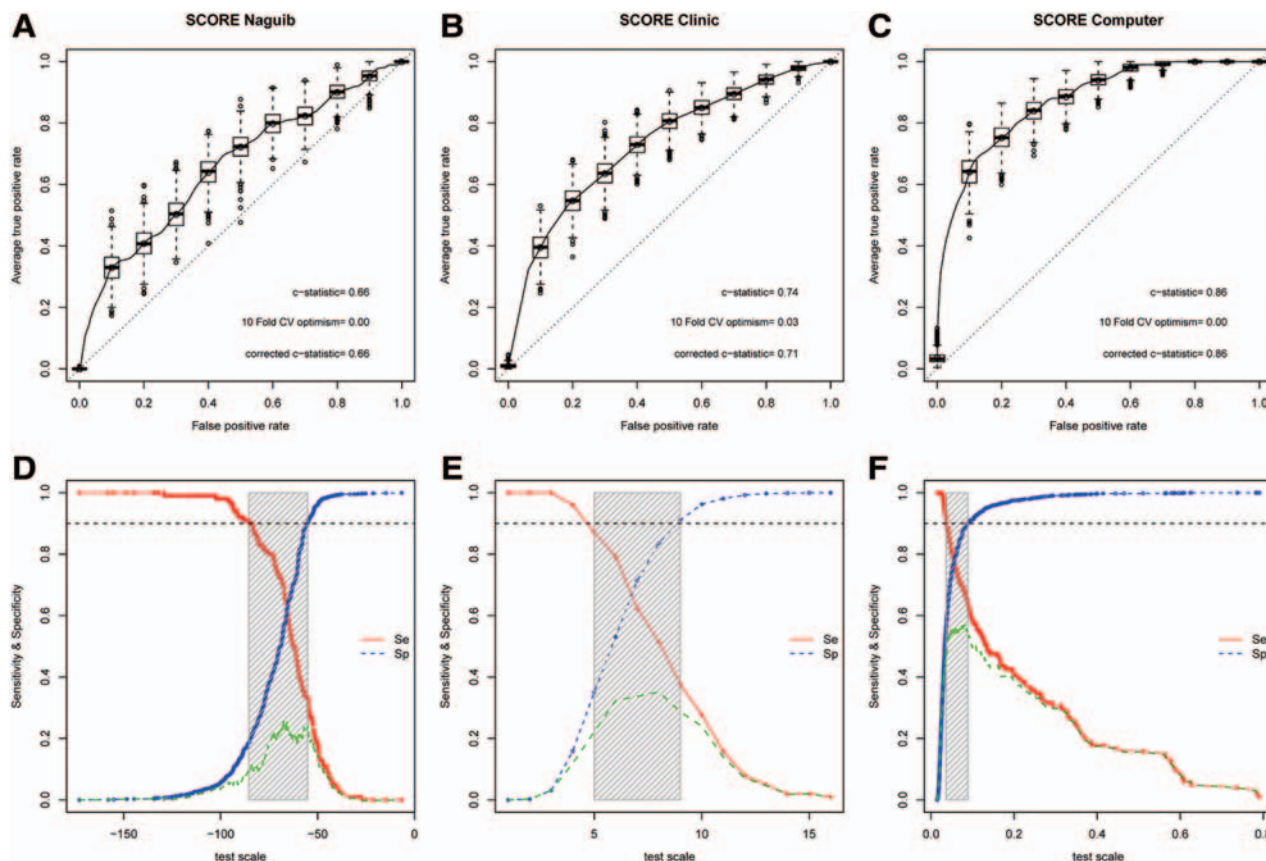


Fig. 3. Average receiver operating characteristic (ROC) curves obtained by 1,000-fold resampling of the study population ($n = 1,359$) for SCORE_{Naguib} (A), the SCORE_{Clinic} (B), and the SCORE_{Computer} (C) ($n = 1,655$); see Methods for explanation of the scores. Two graphs ROC curves with determination of the inconclusive zone (dashed area) using a diagnosis tolerance of 10% for SCORE_{Naguib} (D), the SCORE_{Clinic} (E), and the SCORE_{Computer} (F). AUC_{ROC} = area under the ROC curve; P values refer to difference versus the no discrimination. All AUC_{ROC} were significantly different from 0.5 ($P < 0.001$). CV = cross validation; Se = sensitivity; Sp = specificity.

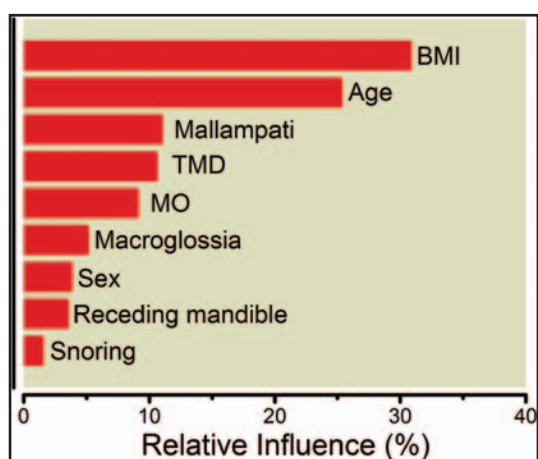


Fig. 4. Illustration of the relative influence of each variable (i.e., the reduction of squared error attributable to each variable) used in the SCORE_{Computer}. BMI = body mass index; MO = mouth opening; TMD = thyromental distance.

which assume, in most of the cases, both an extensive simplification of the complex relationship between continuous variables and outcome, and disregard for the interactions

between variables. In our study a more sophisticated model enables us to improve markedly both the global predictive property as shown by an increase in the area under the ROC curve (fig. 2) and a decrease in the proportion of patients in the inconclusive zone (fig. 3). These results are new data that provide strong arguments to support the hypothesis that individual variables are associated with complex interaction and/or relationship with the outcome and thus that their respective proportions might be as important as their respective values themselves. This can be illustrated by the example of a small patient who has a small MO and a small TMD and is, however, easy to intubate. Height is not recognized as a risk factor for difficult intubation; whereas some of the individual variables are obviously linked to the patient's height.

A promising option for clinical discrimination is to avoid providing a single cutoff that dichotomizes the population, but rather to propose two cutoffs separated by an inconclusive zone, so-called the gray zone.^{13,32} The first cutoff is chosen to exclude the diagnosis with near-certainty whereas the second is chosen to include the diagnosis with near-certainty. This approach is more widely used and results in less loss of information and less distortion than choosing a single

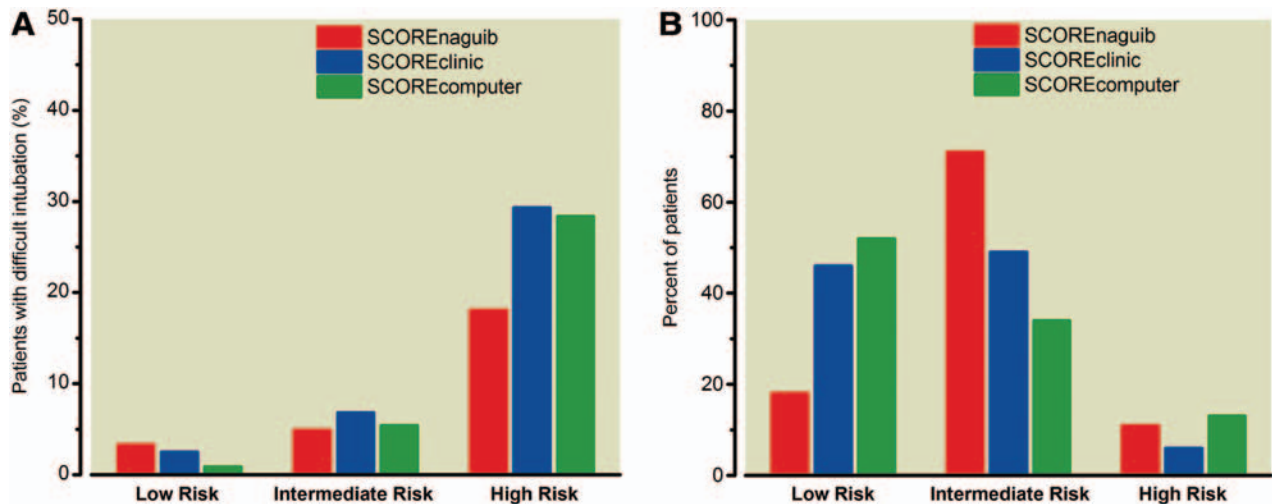


Fig. 5. Percentage of difficult tracheal intubation observed according to the SCORE_{Naguib}, SCORE_{Clinic}, and SCORE_{Computer} score (A) (n = 1,655); see Methods for explanation of the scores. The low-, intermediate-, and high-risk groups were defined using the inconclusive zone, which corresponds to the intermediate-risk group. Proportion of patients in each risk group is provided for the three scores (B).

cutoff, allowing the clinician to improve clinical decision making. This might be particularly important when the ROC curve is relatively flat and thus when it is difficult to choose a single cutoff, which was the case for most of our individual variables (fig. 2). Our study shows that the proportion of patients in the inconclusive zone was important for each individual variable (table 2) and even for most scores, explaining why most of them fail to be considered as accurate predictors of difficult tracheal intubation. Only the SCORE_{Computer} was associated with a highly significant decrease in this proportion (fig. 3). Therefore, we strongly suggest that it is important to assess the risk of difficult tracheal intubation beyond a dichotomous approach, the population being divided into three categories at low, intermediate, and high risk for difficult tracheal intubation to implement an airway-management strategy according to this identified risk. Moreover, a useful score should not only provide appropriate delineation of these three risk categories but also minimize the proportion of patients in the intermediate-risk category.

Some limitations in our study deserve consideration. First, this score was built in an adult population and thus may not apply to pediatric patients. Second, some variables were not taken into account in the predictive model, because of insufficient power because they concerned few patients. This is probably the case of radiotherapy sequelae or burns that modify the skin properties and airway abnormalities that may modify the anatomy of the larynx. We also did not consider a history of previous DTI, which has been identified as an independent risk factor.³³ Nevertheless, because this information is usually either reported or ignored by the patient, documented information may be only incompletely retrieved. Moreover, we think that tracheal intubation should be considered as undoubtedly difficult when such event is reported and no further attempt to predict it

is warranted. Third, we did not assess the reproducibility of measurements of each variables and thus of the different scores. The reproducibility may be limited by significant interobserver variability, which can be expected to worsen in some clinical situations such as emergency. Fourth, boosting approach is one of the best modeling approaches ever developed, which has shown considerable success in predictive accuracy. Its strength is that it requires neither specification of the nature of the interactions between the variables nor hypotheses regarding the linearity of the included variables. However, we do not want to promote boosting approach *per se* but only to suggest that accurate prediction of DTI should take into account interactions and nonlinearities, whatever is the statistical tool used. Last, further studies should be performed in other countries to externally validate our scores.³⁴ However, it should be noted that our study was mainly devoted to a better understanding of the precise mechanisms involved in the prediction of difficult intubation and to open a new paradigm rather than proposing a definite scoring system. Overall, our cohort seemed to behave like most of previously described populations⁹ and the value observed of crossvalidation optimism was low, suggesting that our results are robust. Nevertheless, because our model was data driven we must recognize that an external validation remains mandatory before its clinical application.

In conclusion, although each individual sign of difficult intubation has a poor predictive value, computer-assisted models using complex interaction among simple variables (BMI, MO, TMD, Mallampati class, the presence of receding mandible) enables an accurate prediction of difficult tracheal intubation with a low proportion of patients in the inconclusive zone. This paradigm change in prediction of difficult tracheal intubation might concern other important predictions such as difficult mask ventilation.^{16,35} Further studies are needed to test this hypothesis.

The authors thank David Baker, M.D., F.R.C.A. (Department of Anesthesiology and Critical Care, CHU Necker-Enfants Malades, Paris, France), for reviewing the article, Hinda Meddahi and Djalila Saichi (Research Technicians, Department of Anesthesiology and Critical Care, CHU Pitié-Salpêtrière, Paris, France) for excellent assistance in database management, and colleagues and nurses from our institutions for their help in that study.

Appendix: Generalized Boosted Regression Modeling

Boosting is the process of iteratively adding basis functions in a greedy fashion so that each additional basis function further reduces the selected loss function. We used an implementation close from Friedman Gradient Boosting Machine.²⁵ This procedure is a part of the “gbm” package for R statistical software available on the CRAN package repository.

To improve the reproducibility of this analysis we would like to share the parameters we have retained to conduct this analysis: (1) the total number of trees to fit was set to 10,000; (2) the maximum depth of variable interactions was set to four, which implies a model with up to four-way interactions; and (3) the shrinkage parameter applied to each tree in the expansion was set to 0.001. The following lines of code estimate the GBM model, reproduce the figure 4 (*i.e.*, relative influence of each variable) and draw a ROC curve of the SCORE_{Computer}.

```
# Load libraries of statistical packages within R
require(DiagnosisMed)
require(gbm)
# Read in your dataset
DATA <- read.csv("yourdataset.csv", header = TRUE)
# run GBM model
SCOREcomputer <- gbm(Your.endpoint ~ variable.1 +
variable.2 + ... + variable.n,
  Data = DATA,
  interaction.depth = 4,
  n.trees = 10,000,
  shrinkage = 0.001,
  keep.data = TRUE,
  verbose = TRUE)
# Draw a figure with relative influence and ROC curve
of the model
par(mfrow = c(1,2))
ROC(DATA$Your.endpoint,SCOREcomputer$fit)
summary(SCOREcomputer)
http://cran.r-project.org. Accessed November 2, 2012.
```

CRAN = Comprehensive R Archive Network (R name of the Foundation for Statistical Computing); GBM = Generalized Boosted Regression Modeling; ROC = Receiver Operating Characteristic

References

- Peterson GN, Domino KB, Caplan RA, Posner KL, Lee LA, Cheney FW: Management of the difficult airway: A closed claims analysis. *ANESTHESIOLOGY* 2005; 103:33–9
- American Society of Anesthesiologists Task Force on Management of the Difficult Airway: Practice guidelines for management of the difficult airway. *ANESTHESIOLOGY* 2003; 98:1269–77
- Langeron O, Bourgain JL, Laccoureye O, Legras A, Orliaguet G: [Difficult airway algorithms and management: Question 5. Société Française d'Anesthésie et de Réanimation]. *Ann Fr Anesth Reanim* 2008; 27:41–5
- Combes X, Jabre P, Margenet A, Merle JC, Leroux B, Dru M, Lecarpentier E, Dhonneur G: Unanticipated difficult airway management in the prehospital emergency setting: Prospective validation of an algorithm. *ANESTHESIOLOGY* 2011; 114:105–10
- Amathieu R, Combes X, Abdi W, Housseini LE, Rezzoug A, Dinca A, Slavov V, Bloc S, Dhonneur G: An algorithm for difficult airway management, modified for modern optical devices (Airtraq laryngoscope; LMA CTrach™): A 2-year prospective validation in patients for elective abdominal, gynecologic, and thyroid surgery. *ANESTHESIOLOGY* 2011; 114:25–33
- Isono S, Ishikawa T: Oxygenation, not intubation, does matter. *ANESTHESIOLOGY* 2011; 114:7–9
- Yentis SM: Predicting difficult intubation—worthwhile exercise or pointless ritual? *Anaesthesia* 2002; 57:105–9
- Lundström LH, Møller AM, Rosenstock C, Astrup G, Wetterslev J: High body mass index is a weak predictor for difficult and failed tracheal intubation: A cohort study of 91,332 consecutive patients scheduled for direct laryngoscopy registered in the Danish Anesthesia Database. *ANESTHESIOLOGY* 2009; 110:266–74
- 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
- Wilson ME, Spiegelhalter D, Robertson JA, Lesser P: Predicting difficult intubation. *Br J Anaesth* 1988; 61:211–6
- Arné J, Descoins P, Fuscuardi J, Ingrand P, Ferrier B, Boudigues D, Ariès J: 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, Scamman FL, O'Sullivan C, Aker J, Ross AF, Kosmach S, Ensor JE: Predictive performance of three multivariate difficult tracheal intubation models: A double-blind, case-controlled study. *Anesth Analg* 2006; 102:818–24
- Ray P, Le Manach Y, Riou B, Houle TT: Statistical evaluation of a biomarker. *ANESTHESIOLOGY* 2010; 112:1023–40
- Samsoon GL, Young JR: Difficult tracheal intubation: A retrospective study. *Anaesthesia* 1987; 42:487–90
- 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
- Langeron O, Mazzo E, Huraux C, Guggiari M, Bianchi A, Coriat P, Riou B: Prediction of difficult mask ventilation. *ANESTHESIOLOGY* 2000; 92:1229–36
- Cormack RS, Lehane J: Difficult tracheal intubation in obstetrics. *Anaesthesia* 1984; 39:1105–11
- Horton WA, Fahy L, Charters P: Defining a standard intubating position using “angle finder”. *Br J Anaesth* 1989; 62:6–12
- Amour J, Marmion F, Birenbaum A, Nicolas-Robin A, Coriat P, Riou B, Langeron O: Comparison of plastic single-use and metal reusable laryngoscope blades for orotracheal

- intubation during rapid sequence induction of anesthesia. *ANESTHESIOLOGY* 2006; 104:60–4
20. Amour J, Le Manach YL, Borel M, Lenfant F, Nicolas-Robin A, Carillion A, Ripart J, Riou B, Langeron O: Comparison of single-use and reusable metal laryngoscope blades for orotracheal intubation during rapid sequence induction of anesthesia: A multicenter cluster randomized study. *ANESTHESIOLOGY* 2010; 112:325–32
 21. Youden WJ: Index for rating diagnostic tests. *Cancer* 1950; 3:32–5
 22. Sartorius D, Le Manach Y, David JS, Rancurel E, Smail N, Thicoipé M, Wiel E, Ricard-Hibon A, Berthier F, Gueugniaud PY, Riou B: Mechanism, glasgow coma scale, age, and arterial pressure (MGAP): A new simple prehospital triage score to predict mortality in trauma patients. *Crit Care Med* 2010; 38:831–7
 23. Friedman JH: Greedy function approximation: A gradient boosting machine. *Ann Stat* 2001; 29:1189–232
 24. Friedman J, Hastie T, Tibshirani R: Additive logistic regression: A statistical view of boosting. *Ann Stat* 2000; 28:337–407
 25. Molinaro AM, Simon R, Pfeiffer RM: Prediction error estimation: A comparison of resampling methods. *Bioinformatics* 2005; 21:3301–7
 26. Sullivan LM, Massaro JM, D'Agostino RB Sr: Presentation of multivariate data for clinical use: The Framingham Study risk score functions. *Stat Med* 2004; 23:1631–60
 27. 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
 28. Sing T, Sander O, Beerenwinkel N, Lengauer T: ROCr: Visualizing classifier performance in R. *Bioinformatics* 2005; 21:3940–1
 29. Turgeon AF, Nicole PC, Trépanier CA, Marcoux S, Lessard MR: Cricoid pressure does not increase the rate of failed intubation by direct laryngoscopy in adults. *ANESTHESIOLOGY* 2005; 102:315–9
 30. Martin LD, Mhyre JM, Shanks AM, Tremper KK, Kheterpal S: 3,423 emergency tracheal intubations at a university hospital: Airway outcomes and complications. *ANESTHESIOLOGY* 2011; 114:42–8
 31. 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
 32. Cannesson M, Le Manach Y, Hofer CK, Goarin JP, Lehot JJ, Vallet B, Tavernier B: Assessing the diagnostic accuracy of pulse pressure variations for the prediction of fluid responsiveness: A “gray zone” approach. *ANESTHESIOLOGY* 2011; 115:231–41
 33. Lundström LH, Møller AM, Rosenstock C, Astrup G, Gätke MR, Wetterslev J; Danish Anaesthesia Database: A documented previous difficult tracheal intubation as a prognostic test for a subsequent difficult tracheal intubation in adults. *Anaesthesia* 2009; 64:1081–8
 34. Bleeker SE, Moll HA, Steyerberg EW, Donders AR, Derksen-Lubsen G, Grobbee DE, Moons KG: External validation is necessary in prediction research: a clinical example. *J Clin Epidemiol* 2003; 56:826–32
 35. Kheterpal S, Martin L, Shanks AM, Tremper KK: Prediction and outcomes of impossible mask ventilation: A review of 50,000 anesthetics. *ANESTHESIOLOGY* 2009; 110:891–7