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 (SCOREClinic) and compared it with the more recently described score in the literature (SCORENaguib). Then they used a boosted tree analysis to build the best score possible using computer-assisted calculation (SCORECompur).

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 SCOREClinic model (0.74, 95% CI: 0.72–0.76) was greater than that of the SCORENaguib (0.66, 95% CI: 0.60–0.72, P < 0.001) but significantly lower than that of the SCORECompur (0.86, 95% CI: 0.84–0.91, P < 0.001). The proportion of patients in the inconclusive zone was 71% using SCORENaguib, 56% using SCOREClinic, and only 32% using SCORECompur (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, related to improved airway management through adoption of guidelines with airway management algorithms, widespread use of devices such as gum elastic bougie, laryngeal mask, fiberoptic intubation, new optical devices, and recommendation to prioritize oxygenation. 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.

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). However,
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 sign 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 (video laryngoscope, 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 Samsoon 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 Samsoon 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 oximetry, 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...
was considered as the best score previously described in the literature:

\[ I = 0.2262 - 0.4621 \times \text{TMD} + 2.5516 \times \text{Mallampati score} - 1.1461 \times \text{MO} + 0.0433 \times \text{height} \]

Intubation would be easy if the numerical value (I) in the equation is less than zero (i.e., negative) but difficult if the numerical value (I) is greater than zero (i.e., positive). Some previous studies have provided better predictive scores but they included as a risk factor the fact that the patients already experienced difficult intubation, 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 (SCORE\text{Clinic}), as previously described. 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 (SCORE\text{comp}). 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. 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 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. 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

<table>
<thead>
<tr>
<th></th>
<th>Patients without DTI</th>
<th>Patients with DTI</th>
<th>P Value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(n = 1,554)</td>
<td>(n = 101)</td>
<td></td>
</tr>
<tr>
<td>Men</td>
<td>793 (51)</td>
<td>62 (61)</td>
<td>0.051</td>
</tr>
<tr>
<td>Women</td>
<td>761 (49)</td>
<td>39 (39)</td>
<td></td>
</tr>
<tr>
<td>Age, yr (n = 1,554)</td>
<td>56 ± 17</td>
<td>58 ± 14</td>
<td>0.15</td>
</tr>
<tr>
<td>Age ≥ 75 yr</td>
<td>223 (14)</td>
<td>8 (8)</td>
<td>0.07</td>
</tr>
<tr>
<td>Height, cm</td>
<td>169 ± 10</td>
<td>170 ± 10</td>
<td>0.28</td>
</tr>
<tr>
<td>Weight, kg</td>
<td>75 ± 18</td>
<td>79 ± 19</td>
<td>0.02</td>
</tr>
<tr>
<td>BMI, kg/m²</td>
<td>26.3 ± 5.6</td>
<td>27.5 ± 6.1</td>
<td>0.049</td>
</tr>
<tr>
<td>BMI ≥ 30, kg/m²</td>
<td>310 (20)</td>
<td>29 (29)</td>
<td>0.04</td>
</tr>
<tr>
<td>MO, mm</td>
<td>43 ± 10</td>
<td>38 ± 9</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>TDM, mm</td>
<td>75 ± 18</td>
<td>71 ± 18</td>
<td>0.02</td>
</tr>
<tr>
<td>Mallampati class</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>873 (56)</td>
<td>25 (25)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>2</td>
<td>519 (33)</td>
<td>41 (41)</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>138 (9)</td>
<td>25 (25)</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>24 (1)</td>
<td>10 (10)</td>
<td></td>
</tr>
<tr>
<td>Receding mandible</td>
<td>262 (17)</td>
<td>35 (35)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>History of snoring</td>
<td>145 (9)</td>
<td>18 (18)</td>
<td>0.009</td>
</tr>
<tr>
<td>Macroglossia</td>
<td>62 (4)</td>
<td>14 (14)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Lack of teeth</td>
<td>186 (12)</td>
<td>5 (5)</td>
<td>0.035</td>
</tr>
<tr>
<td>Beard</td>
<td>45 (3)</td>
<td>4 (4)</td>
<td>0.54</td>
</tr>
<tr>
<td>Sellick maneuver</td>
<td>116 (7)</td>
<td>11 (11)</td>
<td>0.24</td>
</tr>
<tr>
<td>Muscle relaxant use</td>
<td>1,442 (93)</td>
<td>87 (86)</td>
<td>0.03</td>
</tr>
<tr>
<td>Type of surgery</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Scheduled</td>
<td>1,271 (82)</td>
<td>82 (82)</td>
<td></td>
</tr>
<tr>
<td>Urgent</td>
<td>214 (14)</td>
<td>17 (17)</td>
<td>0.50</td>
</tr>
<tr>
<td>Unknown</td>
<td>58 (4)</td>
<td>2 (2)</td>
<td></td>
</tr>
<tr>
<td>Orthopedic</td>
<td>317 (20)</td>
<td>21 (21)</td>
<td></td>
</tr>
<tr>
<td>Abdominal</td>
<td>257 (16)</td>
<td>8 (8)</td>
<td></td>
</tr>
<tr>
<td>Urological</td>
<td>147 (9)</td>
<td>12 (12)</td>
<td>0.27</td>
</tr>
<tr>
<td>Vascular and cardiac</td>
<td>129 (8)</td>
<td>8 (8)</td>
<td></td>
</tr>
<tr>
<td>Neurosurgery</td>
<td>384 (25)</td>
<td>27 (27)</td>
<td></td>
</tr>
<tr>
<td>Others</td>
<td>178 (11)</td>
<td>17 (17)</td>
<td></td>
</tr>
<tr>
<td>Unknown</td>
<td>142 (9)</td>
<td>8 (8)</td>
<td></td>
</tr>
<tr>
<td>Cormack and Lehane class</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1,211 (78)</td>
<td>2 (2)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>2</td>
<td>232 (15)</td>
<td>9 (9)</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>44 (3)</td>
<td>57 (56)</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>8 (1)</td>
<td>33 (33)</td>
<td></td>
</tr>
<tr>
<td>Unknown</td>
<td>59 (4)</td>
<td>0 (0)</td>
<td></td>
</tr>
<tr>
<td>Number of laryngoscopy*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1,421 (91)</td>
<td>3 (3)</td>
<td>—</td>
</tr>
<tr>
<td>2</td>
<td>133 (8)</td>
<td>46 (45)</td>
<td></td>
</tr>
<tr>
<td>&gt; 2</td>
<td>0 (0)</td>
<td>52 (51)</td>
<td></td>
</tr>
<tr>
<td>Number of individuals who attempt to intubate*</td>
<td>1,492 (96)</td>
<td>35 (35)</td>
<td>0.005</td>
</tr>
<tr>
<td>1</td>
<td>1,158 (75)</td>
<td>63 (62)</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>178 (11)</td>
<td>13 (13)</td>
<td></td>
</tr>
<tr>
<td>&gt; 2</td>
<td>209 (13)</td>
<td>25 (25)</td>
<td></td>
</tr>
</tbody>
</table>

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 cor-
 responding 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 ± SD and median (95% CI) for non-
normally distributed variables, and numbers (percent-
ages) 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 per-
formed 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 alterna-
tive 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 propen-
sities as well as the proportion of patients in and out of the
inconclusive zone for each main variable (table 2) previ-
ously 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.

---

**PATIENTS INCLUDED**

(n=2,048)

---

**PATIENTS EXCLUDED**

Missing values (n=240)
Plastic blades (n=146)
Alternative techniques first (n=5)
Age < 18 years (n=2)

---

**POPULATION ANALYZED**

(n=1,655)

**Easy intubation**

(n=1,554)

**Difficult intubation**

(n=101)

---

Fig. 1. Flow chart of the study.

**The R Project Contributors. The R project for statistical computing.**

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/m$^2$), 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 SCOREClinic 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 SCOREClinic (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 SCOREClinic compared with the model with continuous variables.

The global predictive ability of the SCOREClinic using ROC curve analysis (0.74, 95% CI: 0.72–0.76) was significantly greater than that of the SCORENaguib (0.66, 95% CI: 0.60–0.72, $P < 0.001$) but significantly lower than that of the SCOREComputer (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 SCOREComputer. 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).

Table 2. Proportion of Patients in the Inconclusive Zone

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

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

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.
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 SCOREClinic in the inconclusive zone, whereas 71% of them had inconclusive values of the SCORENaguib ($P < 0.001$) and 32%, using the SCOREComputer ($P < 0.001$).

### 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,\(^5\),\(^7\) 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,

### Table 3. Multivariate Analysis of Predictors of Difficult Tracheal Intubations and Construction of the SCOREClinic (n = 1,655)

<table>
<thead>
<tr>
<th>Variables</th>
<th>Predictors Analyzed as Continuous Variables</th>
<th>Predictors Analyzed as Discontinuous Variables</th>
<th>Points of the SCOREClinic</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Odds Ratio (95% CI) $P$ Value</td>
<td>Intervals</td>
<td>Odds Ratio (95% CI)</td>
</tr>
<tr>
<td>BMI, kg/m(^2)</td>
<td>1.03 (1.00–1.06) $P = 0.05$</td>
<td>&lt; 25</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>25–34</td>
<td>0.95 (0.65–1.49)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>≥ 35</td>
<td>1.60 (0.76–3.15)</td>
</tr>
<tr>
<td>Mallampati class</td>
<td>2.21 (1.74–2.82) $P &lt; 0.001$</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2</td>
<td>2.58 (1.55–4.38)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3</td>
<td>4.96 (2.67–9.78)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4</td>
<td>7.32 (2.84–18.02)</td>
</tr>
<tr>
<td>MO, mm</td>
<td>0.97 (0.94–0.99) $P = 0.05$</td>
<td>&gt; 50</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>36–50</td>
<td>1.39 (0.66–3.50)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>≤ 35</td>
<td>3.65 (1.59–9.71)</td>
</tr>
<tr>
<td>TMD, mm</td>
<td>0.98 (0.97–1.00) $P = 0.05$</td>
<td>≥ 90</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>61–89</td>
<td>1.24 (0.63–2.66)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>≤ 60</td>
<td>2.04 (1.30–3.34)</td>
</tr>
<tr>
<td>Sex</td>
<td>1</td>
<td>Woman</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>1.52 (1.01–2.37) $P = 0.04$</td>
<td>Man</td>
<td>1.53 (1.01–2.38)</td>
</tr>
<tr>
<td>Receding mandible</td>
<td>1</td>
<td>No</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>2.33 (1.47–3.65) $P &lt; 0.001$</td>
<td>Yes</td>
<td>2.37 (1.49–3.72)</td>
</tr>
</tbody>
</table>

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.

---

Langeron et al.

Anesthesiology 2012; 117:1223-33
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 TDM 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.\textsuperscript{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. 3. Average receiver operating characteristic (ROC) curves obtained by 1,000-fold resampling of the study population (n = 1,359) for SCORE\textsubscript{Naguib} (A), the SCORE\textsubscript{Clinic} (B), and the SCORE\textsubscript{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\textsubscript{Naguib} (D), the SCORE\textsubscript{Clinic} (E), and the SCORE\textsubscript{Computer} (F). AUC\textsubscript{ROC} = area under the ROC curve; P values refer to difference versus the no discrimination. All AUC\textsubscript{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\textsubscript{Computer}. BMI = body mass index; MO = mouth opening; TMD = thyromental distance.
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 

\[\text{SCORE}_{\text{Computer}}\] 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. 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. 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 populations 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. Further studies are needed to test this hypothesis.
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. 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 model, reproduce the figure 4.

```R
# 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)
```

References

intubation during rapid sequence induction of anesthesia. ANESTHESIOLOGY 2006; 104:60–4