Anesthesiology
81:69–75, 1994

J. B. Lippincott Company, Philadelphia

What Is the Best Way to Determine Oropharyngeal Classification and Mandibular Space Length to Predict Difficult Laryngoscopy?

Marc Lewis, M.D., Shahin Keramati, M.D.,* Jonathan L. Benumof, M.D.,†, Charles C. Berry, Ph.D.‡*

Background: **Previous studies have suggested that the degree of visibility of oropharyngeal structures (OP class) and mandibular space (MS) length can predict difficult laryngoscopy. However, those studies were either inconsistent or omit description of how to perform these tests with regard to body, head and tongue position, and the use of phonation, hyoid *versus* thyroid cartilage and inside *versus* outside of the mentum. The purpose of this investigation was to determine which method of testing best predicts difficult laryngoscopy.**

Methods: **In each of 213 consenting adults the OP class was determined in 24 method combinations: two body positions (sitting and supine), three head positions (neutral, sniff, and full extension), two tongue positions (in and out), and with and without phonation. In each patient MS length was measured in 24 method combinations: two body positions (sitting and supine), three head positions (neutral, sniff, and full extension), two distal end points (hyoid and thyroid cartilage), and two proximal end points (inside and outside of the mentum). In each patient the laryngoscopic grade was determined at the time of induction of anesthesia. We defined laryngoscopic grades III (n = 24) and 4 (n = 0) as difficult. The area under the receiver operating characteristic curve (ROC area) for each combination was used to compare the combinations and determine significant differences: ROC area = 0.5 implied a totally uninformative combination and ROC area = 1.0 a combination that predicted perfectly. Logistic regression analysis was used to calculate a predictor of difficult intubation that combined both OP class and MS length (the performance index). The performance index could then be used to calculate sensitivity, specificity, positive and negative predictive value, and probability of difficult intubation.**

Results: **The ROC areas for the different combinations used to assess OP class ranged from 0.78 to 0.94. The best combination was with the patient sitting, head in extension, tongue out, and with or without phonation. For MS length, the ROC areas ranged from 0.58 to 0.77; the best combination was the patient sitting, with the head in extension, with distance measured from the inside of the mentum to the thyroid cartilage. Combining the OP class and MS length (performance index = 2.5 × OP class − MS length in centimeters) significantly increased predictability of difficult intubation. At performance index = 0 and = 2, the probability of difficult intubation was 3.5% and 24%, respectively. With clinically relevant cutpoints for the performance index it was found that most difficult intubations could be predicted, but approximately half of those predicted to be difficult would in fact be easy.**

Conclusions: **Based on the above ROC areas and ease of performing the test for the patient, we recommend that these tests be performed with patients in the sitting position, with the head in full extension, the tongue out, and with phonation, and with distance measured from the thyroid cartilage to inside of the mentum. Nevertheless, it is clear that these two tests, either used alone or in combination, will fail to predict a few difficult laryngoscopies and that they will predict difficult laryngoscopy in a significant number of patients in whom the trachea is easy to intubate. (Key words: Airway management. Laryngoscopy. Mandibular space. Oropharyngeal classification. Preoperative evaluation. Tracheal intubation.)**

PREVIOUS studies have suggested that the visibility of the oropharyngeal structures (OP class) and measurement of the length of the mandibular space (MS) can predict difficult laryngoscopy. However, those studies were either inconsistent or omit description of how to perform these tests with regard to body and head positions for both tests as well as the use of phonation and tongue position for the OP class, and the use of the hyoid or thyroid cartilage and the inside (posterior surface) or outside (anterior surface) of the mentum (most anterior part of the mandible) for the MS length. Each of these measurement method combinations might have a major impact on the result of the measurement.

With respect to the MS length, Freck[1] made the measurement with the head position in full extension, whereas Roucke *et al.*[2] had their patients in a neutral head position. In addition, Roucke *et al.*[2] did not spec-

* Medical Student IV.

† Professor of Anesthesia.

‡ Adjunct Professor of Family and Preventive Medicine.

Received from the Department of Anesthesiology, University California, San Diego Medical Center. Accepted for publication March 18, 1994.

Address reprint requests to Dr. Benumof: University of California, San Diego Medical Center, Mail Code 0801, 402 Dickinson Street, San Diego, California 92103-0801.

ify whether the proximal end point was the inside or outside of the mentum. Finally, Cohen *et al.*[3] used both the sitting and supine body positions but did not specify which position was used for any given patient; therefore, the body position is unknown. With respect to OP class, Mallampati *et al.*[4] and Samsoon and Young[5] omitted mention of the use of phonation, whereas Freck[1] and Roucke *et al.*[2] specified that phonation was not used and Cohen *et al.*[3] specified that phonation was used. When analyzing OP class, Freck[1] and Cohen *et al.*[3] did not specify the patient's head position, but Roucke *et al.*,[2] Samsoon and Young,[5] and Tham *et al.*[6] used the neutral head position, and Mallampati *et al.*[4] used the sniff position.

It is clear from the forgoing that there is no standard method of performance of the OP class and MS length tests. Consequently, the results of these tests are difficult to interpret from patient to patient and from study to study. The reason no standard method of performance exists is that it is not known which method of performance best predicts the difficulty (grade) of laryngoscopy. The purpose of this investigation was to measure every possible measurement combination (n = 24 for each test) and to correlate the result of each measurement combination with the laryngoscopic grade of laryngeal exposure obtained at time of induction of anesthesia. From these correlations we determined which OP class and MS length methods best predicted a difficult laryngoscopy.

Materials and Methods

The study population of this institutionally approved study consisted of 213 individually consenting adults (135 men and 78 women; age 41 ± 11 yr [mean ± SD], range 19–80 yr; weight 80 ± 16 kg, range 48–137 kg; height 170 ± 15 cm, range 152–188 cm) who had no obvious anatomic or pathologic abnormality of the head and/or neck and who required general anesthesia and endotracheal intubation. In each patient, OP class [with respect to visualization: 1 = soft palate, fauces, uvula, anterior and posterior tonsillar pillars; 2 = soft palate, fauces, uvula; 3 = soft palate, base of uvula; 4 = soft palate (uvula not visible at all)][4,5,7] was determined preoperatively by inspection (aided by a bright pen light) in 24 different method combinations: two body positions (sitting and supine), three head positions (neutral, sniff, and full extension), two tongue positions (in and out), and with and without phonation ("ahhh").

In each patient, MS midline length was measured preoperatively (in millimeters by caliper) in 24 different method combinations: two body positions (sitting and supine), three head positions (neutral sniff and full extension), two distal end points (hyoid and thyroid cartilage), and two proximal end points (inside and outside the mentum). In each patient the best possible laryngoscopic grade (with respect to visualization: 1 = vocal cords; 2 = posterior portion of the laryngeal aperture; 3 = the epiglottis and soft palate; 4 = only the soft palate)[5,7,8] was determined by an experienced laryngoscopist (minimum experience was 2 yr) using a Macintosh laryngoscope blade, with all patients anesthetized, fully paralyzed (documented by blockade monitor), and in the sniff position. The sniff position for the determination of OP class, measurement of MS length and determination of laryngoscopic grade was obtained by using a premolded head piece (model 40-0075, invented by Popitz M, American Medical Development, Boston, MA).

All preoperative measurement combinations in all patients were evaluated by one of two different observers (ML and SK) in the following sequence. The patients were first supine with their heads in a neutral position. OP class was measured with the tongue in, and then out, both with and without phonation. The MS length was then measured without moving the body or head position, from two distal end points (the hyoid and thyroid cartilage) to each the inner mentum and the outer mentum. This sequence of measurements was repeated, with the patient remaining in the supine position, both with the head in the sniff position and in full extension (head maximally extended on the neck). The patient was then placed in a sitting position, at 80 degrees, and the sequence of measurements was repeated as above, for both the OP class and the MS length with the head in the neutral, sniff, and fully extended positions. Thus a total of 48 measurements were recorded for each patient (24 combinations for the OP class and 24 combinations for the MS length). Each laryngoscopist was unaware of these study findings, but each laryngoscopist did whatever independent preoperative evaluation they thought was consistent with standard of care.

Statistical Analysis

Table 1 defines all statistical terms used in the data analysis and presentation of results. We defined laryngoscopic grades 3 (n = 24) and 4 (n = 0) as "difficult." The method of receiver operating characteristic (ROC)

Table 1. Definition of Statistical Terms Used in Data Analysis and Presentation of Results

Statistical Term		Definition
Sensitivity	=	Number of difficult intubations correctly predicted / Number of difficult intubations
Specificity	=	Number of easy intubations correctly predicted / Number of easy intubations
True-positive fraction	=	Sensitivity
False-positive fraction	=	1 − specificity
Positive predictive value	=	Number of difficult intubations correctly predicted / Number of intubations predicted to be difficult
Negative predictive value	=	Number of easy intubations correctly predicted / Number of intubations predicted to be easy

analysis was used to compare the predictive value of different combinations for determining OP class and for measuring MS length.[9] The ROC curve is formed by plotting the true positive fractions (sensitivity) *versus* the false positive fractions (1 − specificity), and the ROC area is the area under this curve. The ROC area ranges from 0.5 (corresponding to a totally uninformative variable) to 1.0 (corresponding to a variable which classifies perfectly).

The area under the ROC curve was used as the criterion to compare the different combinations. Statistical comparisons of ROC curves are based on the methods of DeLong *et al.*[10] Using suitable linear contrasts of the ROC areas (*e.g.*, differences of means) for the different combinations, tests were constructed of the hypotheses that the average of all areas exceeds 0.5, that the areas differ (by more than chance variation), and that specified combinations achieve higher areas than others.

Logistic regression was used to calculate a predictor of difficult intubation that combined both OP class and MS length (the performance index). The OP class determined without phonation, tongue out, sitting, with head in extension and the MS length determined using thyromental distance, inside mentum, sitting, with head in extension were used for the logistic regression. The performance index was then used to calculate sensitivity, specificity, positive and negative predictive value and probability of difficult intubation.

The performance index can be used to predict whether a given patient's trachea will be difficult to intubate by classifying the intubation as "difficult" if that patient's performance index exceeds a certain value (called a "cutpoint") and "not difficult" if the index does not exceed that value. It is important to select this cutpoint so that a balance is found between having false positive cases (which will be more numerous when the cutpoint is set to a low value) and having false negative cases (which will occur more often when the cutpoint is set to a high value).

Results

Of 213 patients, in 24 laryngoscopy was difficult (grade 3). The ROC areas for the different combinations used to assess OP class ranged from 0.78 to 0.94 and are displayed in Table 2. Each of these 24 combinations attains $P < 0.01$ even when the Bonferroni correction is applied, indicating that the ROC areas are greater than would be observed by chance alone. The predictive value of these varies among the combinations ($P < 0.0001$). The average ROC area with phonation was 0.048 greater than that without phonation ($P < 0.0001$). The average area with the tongue out was 0.043 greater ($P < 0.0001$) than with tongue in. The average area while sitting was 0.027 greater than that while supine ($P < 0.0001$). The average area was 0.029 greater with the head in extension than in sniff position ($F = 0.0006$) and 0.011 greater in the sniff position than in the neutral position ($P = 0.20$).

The combination formed by using the better (or best) of each of these (*i.e.*, with phonation, tongue out, sitting, and head in extension) achieved a ROC area of 0.937. The greatest ROC area of 0.938 was achieved by the combination which is identical except for phonation. It is not possible to determine whether these two combinations actually differ (the attained P value is 0.92); it seems that either would give good results for determining OP class.

For MS length, the ROC areas ranged from 0.58 to 0.77 (table 3). All but one attains $P < 0.05$ and 15

Table 2. Receiver Operating Characteristic Areas for Oropharyngeal Class

			Head Position		
Phonation	Tongue	Body Position	Neutral	Sniff	Extension
No	In	Supine	0.78	0.80	0.84
Yes	In	Supine	0.88	0.89	0.91
No	Out	Supine	0.88	0.86	0.90
Yes	Out	Supine	0.91	0.91	0.93
No	In	Sitting	0.80	0.84	0.90
Yes	In	Sitting	0.89	0.92	0.93
No	Out	Sitting	0.90	0.91	0.94
Yes	Out	Sitting	0.92	0.92	0.94

Table 3. Receiver Operating Characteristic Areas for Mandibular Space Distance

			Head Position		
Mentum	Distal End Point	Body Position	Neutral	Sniff	Extension
Inside	Hyoid	Supine	0.61	0.73	0.75
Outside	Hyoid	Supine	0.61	0.73	0.72
Inside	Thyroid	Supine	0.65	0.74	0.76
Outside	Thyroid	Supine	0.58	0.69	0.73
Inside	Hyoid	Sitting	0.64	0.74	0.75
Outside	Hyoid	Sitting	0.60	0.71	0.73
Inside	Thyroid	Sitting	0.67	0.76	0.77
Outside	Thyroid	Sitting	0.61	0.71	0.73

combinations still attain $P < 0.05$ when the Bonferroni correction is applied suggesting that MS length predicts difficult laryngoscopy although not as well as OP class. The differences among these ROC areas are not merely due to chance ($P = 0.008$). The extended head position had an average area 0.014 greater than the sniff head position ($P = 0.47$), and the sniff head position had an average area 0.105 greater than the neutral position ($P = 0.0007$). This latter difference is by far the largest. None of the other variations in measuring MS length had a significant effect. The average ROC area for measuring inside the mentum was 0.036 greater than that for measuring outside the mentum ($P = 0.09$). The average was 0.007 greater for thyromental distance than for hyomental distance ($P = 0.58$) and 0.012 greater when sitting than supine ($P = 0.41$).

The combination formed by using the better (or best) of each of these (*i.e.*, using the thyroid cartilage, the inside of the mentum, sitting, and head in extension) achieved a ROC area of 0.767. This area was larger than that of any other combination, but there were several combinations that differed just slightly (and not statistically significantly) from this.

A question of some interest is whether basing the prediction of difficult laryngoscopy on both OP class and MS distance is superior to using just the OP class. Logistic regression analysis was used to find a predictor of difficult laryngoscopy based on a linear combination (*i.e.*, a weighted average) of the OP class and the MS length which maximizes the likelihood of the predictions. Using the best OP class and MS length, that combination (performance index) is 2.5 × OP class − MS length in centimeters). The ROC area for that predictor is 0.967, which improves upon predictions based on either OP class ($P = 0.001$) or MS length ($P < 0.0001$) alone.

Using the result that 2.5 × OP class − MS length in centimeters provides the best index for predicting difficult laryngoscopy, a simple rule for predicting difficult laryngoscopies can be constructed. Figures 1 and 2 plot several measures of diagnostic performance for this index. As seen in figures 1 and 2, at any level of performance index, sensitivity, specificity, positive and negative predictive value, and probability (for various prevalences) of difficult intubation may be determined. For example, in settings in which the prevalence of difficult intubation is 11%, the probability that a patient who has a performance index of exactly zero will in fact be difficult to intubate is 3.5% (fig. 2). When this patient and all patients who have higher values of the performance index are combined into a single group, 39% of the group will actually have a difficult intubation (fig. 1, positive predictive value) and 100% of the patients who will have difficult laryngoscopies are in this group (fig. 1, sensitivity). None of the patients in the other group (performance index < 0) will have a difficult laryngoscopy (fig. 1, negative predictive value) and 80% of all patients in whom the trachea is easy to intubate (grade 1 or 2) fall into this group (fig. 1, specificity). Increasing the cutpoint to 2.0 increases the probability of difficult intubation to 24% (at a prevalence of 11%) and the percentage of grade 1 or 2 laryngoscopies that are predicted to be easy decreases to 93%. The percentage of those predicted to be difficult that actually turn out that way increases to 61%, and the percentage of those predicted not to be difficult that turn out that way is 98%. Thus, predicting an ever-increasing proportion of the difficult laryngoscopies is at the cost of falsely predicting ever more cases in which the laryngoscopy is not actually difficult. Finally, it can be seen from figure 2 that increasing the prevalence from 5% to 15% at a performance index of 2 greatly increases the probability of difficult intubation from just over 5% to over 40%.

Discussion

We found that OP class and MS distance can be used to predict difficult laryngoscopy with a reasonable degree of accuracy. However, the most important clinical implications are that these two tests alone will fail to predict some difficult laryngoscopies and that they will predict difficult laryngoscopy in a significant number of patients who in fact are easy to intubate. Thus, the cost of being prepared for virtually every difficult la-

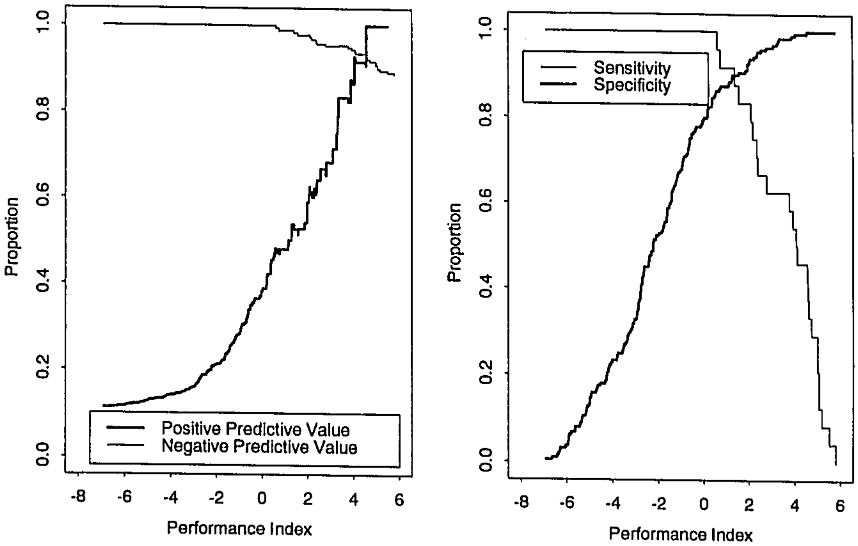

Fig. 1. The x-axis is the performance index, calculated as oropharyngeal class × 2.5 − mandibular space length in centimeters. The figure plots several measures of diagnostic performance (y-axis, *left* and *right*) for a wide range of performance indexes.

ryngoscopy is that in approximately half the cases the intubation will be easy.

These results are not totally unexpected because the two tests are inherently limited: they do not account for all of the factors that make laryngoscopy difficult. Anatomic factors besides relative tongue and pharyngeal size (OP class) and the relative location of the larynx (MS length) that correlate with a difficult laryngoscopy include decreased atlantooccipital joint extension, decreased compliance of tissue (tongue, submandibular space), decreased neck length and increased neck muscularity, overriding maxilla, enlarged incisors, decreased temporal-mandibular joint mobility, and a narrow highly arched palate.

There are several potential limitations of the study related to experimental design. First, there were two different observers who recorded the patient measurements in the 48 different combinations. How-

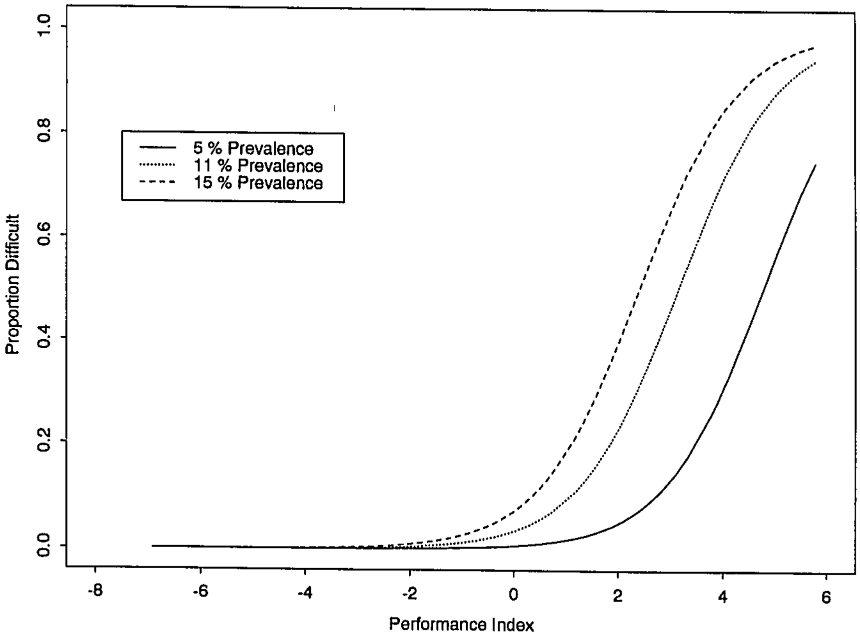

Fig. 2. The x-axis is the performance index, calculated as oropharyngeal class × 2.5 − mandibular space length in centimeters. The y-axis is the probability of difficult intubation as determined by logistic regression analysis. The figure plots this relation as a function of three prevalences of difficult intubation; in this study, the prevalence of difficult intubation was 11%.

ever, because both observers had identical training and used the same methods to make these simple measurements and because the significant ROC areas were high (which implies reproducibility), we do not believe that the use of two observers was a source of significant variation or error. Second, several different anesthesia residents, nurse anesthetists and faculty members intubated the tracheas, but all intubating individuals were experienced at doing so and all obtained the best possible view; we are confident that variation in the laryngoscopic grade due to the experience of the laryngoscopist was of minor importance. Third, the sniff position either during the evaluation, or at the time of intubation, could have varied from patient to patient. However, the head was stabilized in the sniff position by a commercial head and neck mold, in both instances in all patients. Fourth, we found an 11% prevalence of grade 3 laryngoscopies, which is moderately increased compared with that found in the previous studies.[7] Potential reasons for an increased incidence of grade 3 laryngoscopies in our study include the use of multiple laryngoscopists, and use of in-hospital patients (*e.g.*, fluid balance problems, previous airway manipulations, or exposure to chemotherapy or radiation therapy). However, we eliminated any patient with a known obvious pathologic (*e.g.*, mass of any kind or traction) or physiologic impediment (*e.g.*, edema) involving the head or neck from the study. In addition, only the Macintosh laryngoscope blade was used for intubations. Fifth, the reader should be cautioned that estimates of diagnostic accuracy which are based on one set of patients (those that were used to develop the diagnostic rule) tend to overstate the true accuracy. Further, the incidence of difficult laryngoscopy will depend on the nature of the clinical case mix; in settings with lower prevalence a higher performance index cutpoint may be justified. Thus, although we expect the proposed rule to be accurate and clinically useful, prospective monitoring of its performance, and possible adjustments in the recommended cutpoint, are advised.

With respect to determining OP class, we recommend the sitting position with full head extension, with tongue out and phonation for the following reasons. First, the sitting position results were generally more predictive than the supine position. In addition, the test took less time and effort for the patient to produce a nonvarying result in the sitting position. Second, the sniff and full extension head positions are far better than the neutral head position, and full extension is much easier to reproduce than the sniff position without the use of a head stabilizing device. Third, the results are clearly better with patient phonation and the tongue and pharynx are more motionless when the patient phonates. Finally, it is easier for the patient to perform the tests and the results are more reproducible with the tongue out (the tongue was often not motionless in the tongue in position), although tongue out and tongue in results are almost the same.

The MS length was a slightly poorer predictor than OP class. Just as in the case of the OP class, it is preferable to perform the test in the sitting position, although similar results can be achieved in the supine position. Additionally, the sniff and full extension are superior to the neutral head position, with the full extension remaining more reproducible. It seems better to use the inner mentum as a proximal landmark, possibly because there is so much variation in the size of the fat pad on the bony point of the chin (the outer mentum). There was no difference in the results between use of the thyromentum as opposed to the hyomentum as a distal landmark, but the thyomentum seemed much easier to find consistently, especially in obese patients. When both the MS length and OP class were used simultaneously the above conclusions remained unchanged.

The greatest value of these tests may lie in that they require the examiner to think about the laryngoscopic view and an appropriate anesthetic plan in advance of the procedure. In addition, the performance of these tests yields a great deal of additional information about the morphology of the palate, the flexibility, length, and muscularity of the neck, maxillary override on the mandible, the size of the incisors, and the extent to which the patient can open the mouth. Nevertheless, part of the assessment of the airway remains subjective. This study suggests that a more powerful method of predicting difficult laryngoscopy may be developed by including additional tests in the preoperative analysis. In addition, it would be of value to compare the predictive value of these tests performed with a Miller blade as compared with a Macintosh blade.

References

1. Freck CM: Predicting difficult intubation. Anaesthesia 46:1005–1008, 1991

2. Roucke DA, Murray WB, Rout CC, Gouws E: Relative risk analysis of factors associated with difficult intubation in obstetric anesthesia. ANESTHESIOLOGY 77:67–73, 1992

3. Cohen SM, Laurito CE, Segil LJ: Examination of the hypopharynx predicts ease of laryngoscopic visualization and subsequent intubation: A prospective study of 665 patients. J Clin Anesth 4:310–314, 1992

4. Mallampati SR, Gatt SP, Gugino LD, Desai SP, Waraksa B, Freiberger D, Liu PL: A clinical sign to predict difficult intubation: A prospective study. Can J Anaesth 32:429–434, 1985

5. Samsoon GT, Young JB: Difficult tracheal intubation: A retrospective study. Anaesthesia 42:487–490, 1987

6. Tham EJ, Gilderslevc CD, Sanders LD, Mapleson WW, Vaughan RS: Effects of posture, phonation and observer on Mallampati classification. Br J Anaesth 68:32–38, 1992

7. Benumof JL: Management of the difficult adult airway. ANESTHESIOLOGY 75:1087–1110, 1991

8. Cormack RS, Lehane J: Difficult tracheal intubation in obstetrics. Anaesthesia 39:1105–1111, 1984

9. Metz CE: ROC methodology in radiologic imaging. Invest Radiol 21:720–733, 1986

10. DeLong ER, DeLong DM, Clarke-Pearson DL: Comparing the areas under two or more correlated receiver operating characteristic curves: A nonparametric approach. Biometrics 44:837–845, 1988