

Predicting Difficult Intubation in Apparently Normal Patients

A Meta-analysis of Bedside Screening Test Performance

Toshiya Shiga, M.D., Ph.D.,* Zen'ichiro Wajima, M.D., Ph.D.,† Tetsuo Inoue, M.D., Ph.D.,‡ Atsuhiro Sakamoto, M.D., Ph.D.§

The objective of this study was to systematically determine the diagnostic accuracy of bedside tests for predicting difficult intubation in patients with no airway pathology. Thirty-five studies (50,760 patients) were selected from electronic databases. The overall incidence of difficult intubation was 5.8% (95% confidence interval, 4.5–7.5%). Screening tests included the Mallampati oropharyngeal classification, thyromental distance, sternomental distance, mouth opening, and Wilson risk score. Each test yielded poor to moderate sensitivity (20–62%) and moderate to fair specificity (82–97%). The most useful bedside test for prediction was found to be a combination of the Mallampati classification and thyromental distance (positive likelihood ratio, 9.9; 95% confidence interval, 3.1–31.9). Currently available screening tests for difficult intubation have only poor to moderate discriminative power when used alone. Combinations of tests add some incremental diagnostic value in comparison to the value of each test alone. The clinical value of bedside screening tests for predicting difficult intubation remains limited.

UNANTICIPATED difficult intubation can be challenging to anesthesiologists. Numerous investigators have attempted to predict difficult intubation by using a simple bedside physical examination. Mallampati *et al.*¹ introduced in 1985 a currently well-known screening test that classifies visibility of the oropharyngeal structure. The distance from the thyroid notch to the mentum (thyromental distance), the distance from the upper border of the manubrium sterni to the mentum (sternomental distance), and a simple summation of risk factors (Wilson risk sum score) are widely recognized as tools for predicting difficult intubation.^{2,3} Nevertheless, the diagnostic accuracy of these screening tests has varied from trial to trial,⁴ probably because of differences in the incidence of difficult intubation, inadequate statistical power, different test thresholds, or differences in patient characteristics. Questions remain as to whether a com-

bination of tests may improve predictive accuracy or whether predictive accuracy differs for specific groups of patients, such as obstetric or obese patients, in whom difficult intubation is considered to occur more often than in normal patients. A recent editorial by Yentis⁵ made clear how hard it is to predict difficult intubation because of its low rate of occurrence and questioned whether attempts at prediction are likely to be useful.

To answer these questions, we systematically reviewed and synthesized the published data relating to the performance of diagnostic tests for difficult intubation in normal-airway patients scheduled to undergo general anesthesia.

Study Selection and Quality Assessment

We searched MEDLINE (1980 through May 2004) and the Cochrane Central Register of Controlled Trials (2004, issue 1) for reports of studies and trials relating to the accuracy of predictive tests for difficult intubation. No language restrictions were applied. The initial search terms were *difficult airway*, *difficult intubation*, and *difficult laryngoscopy*. A manual search of references cited in published reports and reviews was also performed.

Reports were independently selected and reviewed by two investigators (T. S. and Z. W.). The systematic review process for selection of eligible studies is shown in figure 1. Reported studies were selected if they met the following criteria: (1) the study was prospective; (2) at least one bedside diagnostic test was used; (3) absolute numbers of true-positive, false-negative, true-negative, and false-negative results were available or could be derived from the published data; and (4) a standard laryngoscope was used. We did not include articles from any study with insufficient data, with patients whose airway was anatomically abnormal, or with a complicated and not widely accepted scoring system. We excluded retrospective studies, studies requiring impractical and costly diagnostic tests that are not yet widely accepted (e.g., radiologic examinations), and studies involving a special laryngoscope or technique.

The quality of acceptable reports was assessed independently by two authors (T. S. and Z. W.). Studies were graded according to four *a priori* criteria for quality, as described by Romagnuolo *et al.*⁶: (1) blinding, (2) consecutive recruitment of patients, (3) single (*vs.* composite) accepted standards, and (4) nonselective use of the accepted standard. One point was given for each criterion; the maximum possible score was 4.

This article is featured in "This Month in Anesthesiology."
Please see this issue of Anesthesiology, page 5A.

* Staff Anesthesiologist, † Assistant Professor, ‡ Professor, Nippon Medical School Chiba Hokusoh Hospital. § Professor and Chairman, Nippon Medical School Hospital, Tokyo, Japan.

Received from the Department of Anesthesia, Nippon Medical School Chiba Hokusoh Hospital, Chiba, Japan. Submitted for publication July 30, 2004. Accepted for publication April 1, 2005. Support was provided solely from institutional and/or departmental sources.

Address reprint requests to Dr. Shiga: Department of Anesthesia, Nippon Medical School Chiba Hokusoh Hospital, Kamagari 1715, Inba-mura, Inba-gun, Chiba 270-1694, Japan. Address electronic mail to: qzx02115@nifty.com. Individual article reprints may be purchased through the Journal Web site, www.anesthesiology.org.

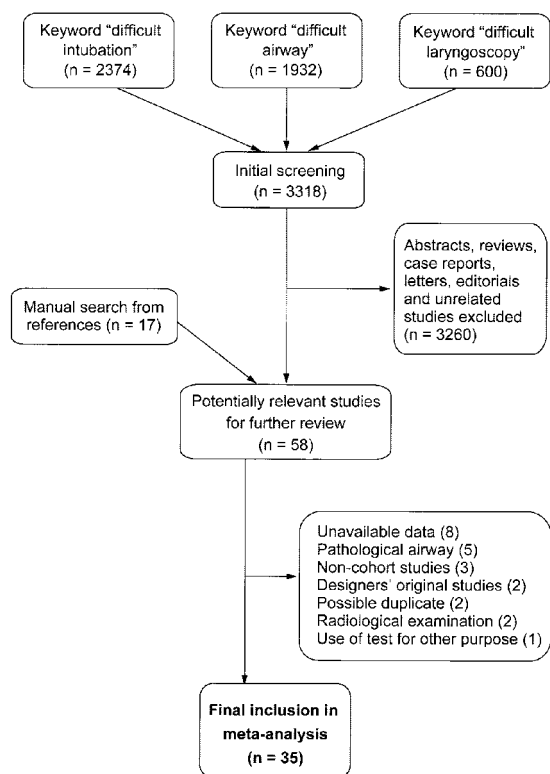


Fig. 1. Meta-analysis flowchart.

Data Extraction

We defined a Cormack-Lehane grade⁷ of 3 or greater as the accepted standard for difficult intubation, as in most studies included in the review. Some authors reported a required special technique, multiple unsuccessful attempts, or a combination of these as the accepted standard for difficult intubation. If a definition seemed so subjective as to be generally unacceptable, we abandoned it and substituted the Cormack-Lehane grades whenever available. Extracted from the reports were the number of patients, mean age, general patient characteristics, criteria for difficult intubation, type of laryngoscope blade, incidence of difficult intubation, type of screening test, and absolute numbers of true-positive, false-positive, true-negative, and false-negative results. For a diagnostic test to be included in our analysis, at least three reports of that test had to have been identified in our literature search. The Mallampati classification, thyromental distance, sternomental distance, mouth opening, Wilson risk sum score, and a combination of the Mallampati classification and thyromental distance met our inclusion criteria. In addition, we performed subgroup analysis. Obstetric or obese patient groups were analyzed if they were grouped separately within a study and the subgroup appeared in at least three studies.

Data Synthesis

We calculated pooled estimates of the incidence of difficult intubation, sensitivity, specificity, positive and

negative likelihood ratios, and natural logarithm of diagnostic odds ratio by the DerSimonian-Laird random-effects model.⁸ Rates were pooled after logit transformation, weighting study rates by the inverse ratio of their variance plus the between-study variance for that measure, and then retransformed back into standard proportions with 95% confidence intervals (CIs). Homogeneity of the effect size across trials was tested by chi-square statistics. Heterogeneity was defined as $P < 0.1$.

Sensitivity is the ratio of the true-positive number to the sum of true-positive plus false-negative numbers. Specificity is the ratio of the true-negative number to the sum of true-negative plus false-positive numbers.⁹ Likelihood ratios are obtained as follows: positive likelihood ratio = sensitivity/(1 - specificity); negative likelihood ratio = (1 - sensitivity)/specificity. Likelihood ratios greater than 10 and less than 0.1 are considered strong evidence for ruling in or ruling out diagnoses, respectively, under most circumstances.¹⁰ The log diagnostic odds ratio is the logit (positive likelihood ratio/negative likelihood ratio), indicating a summary of diagnostic performance.¹¹

The diagnostic performance of each test was also assessed by means of summary receiver operating characteristic (ROC) curves according to the method described by Moses *et al.*¹² We constructed ROC curves. Briefly, the true-positive rate was plotted against the false-positive rate for each study. To avoid calculation problems by having values of zero, 0.5 was added to each cell of the respective contingency table. The summary ROC model is described by the following equation $D = a + bS$. The summary ROC curve analysis is based on regression analysis of logit transformed data, which plots the difference between the logit of the true-positive rate (TPR) and the logit of the false-positive rate (FPR) ($D = \text{logit TPR} - \text{logit FPR}$) on the y-axis and their sum ($S = \text{logit TPR} + \text{logit FPR}$) on the x-axis. The y-axis (D) is equivalent to the log diagnostic odds ratio, and the x-axis (S) is a measure of how the test characteristics vary with the test threshold. The regression coefficient b examines the extent to which the log odds ratio is dependent on the threshold values chosen. The linear regression analysis was weighted by the inverse of the variance of D . The regression line was back-transformed to the ROC space.

Assessment of Publication Bias

To assess the potential for publication bias, a funnel plot was constructed in which the log of relative risks was plotted against the associated number of patients.¹³ In addition, correlation between standardized log relative risks and the associated number of patients was determined by the Kendall rank correlation coefficient. The correlation between sample size and relative risk would be strong if not many small studies with null results were published. A significant correlation between sample size and relative risk would not exist in

the absence of publication bias. Statistical significance was defined for treatment effects as $P < 0.05$ and for heterogeneity and publication bias as $P < 0.1$. Analyses were performed with Microsoft Excel (Microsoft Corporation, Redmond, WA), Meta-DiSc[®] (Hospital Ramón y Cajal, Madrid, Spain), and Number Cruncher Statistical System 2004 (NCSS Statistical Systems, Kaysville, UT).

Results

The electronic search resulted in 3,318 hits. Thirty-five studies¹⁴⁻⁴⁸ representing 50,760 patients met the inclusion criteria (table 1). Non-English-language reports included 3 in French, 2 in German, 1 in Italian, and 1 in Japanese. We excluded the original articles by Mallampati *et al.*¹ and Wilson *et al.*³ because the test designers were also the test assessors. One report⁴⁹ was included in the analysis of obese populations but was excluded from the final analysis because of possible duplication of data.

The overall incidence of difficult intubation was 5.8% (95% CI, 4.5-7.5%) for the overall patient population, 6.2% (95% CI, 4.6-8.3%) for normal patients excluding obstetric and obese patients, 3.1% (95% CI, 1.7-5.5%) for obstetric patients, and 15.8% (95% CI, 14.3-17.5%) for obese patients. Data pertaining only to obstetric patients were given in four reports. Of these reports, three assessed risk on the basis of the Mallampati classification and one assessed risk on the basis of sternal distance; therefore, we analyzed only the Mallampati test data. Data pertaining exclusively to obese patients were given in four reports, all of which assessed risk on the basis of the Mallampati test.

Pooled estimates of the incidence of difficult intubation, sensitivity, specificity, positive and negative likelihood ratios, and natural logarithm of diagnostic odds ratio as well as the regression model equation for each test are shown in table 2. The summary ROC curve for each test is shown in figure 2. With the exception of thyromental distance, diagnostic accuracy did not vary with the test threshold in any test. Because diagnostic accuracy tended to vary with the test threshold for thyromental distance ($P = 0.056$), we calculated likelihood ratios with an adjusted cutoff point; a stricter criterion for thyromental distance (≤ 6.0 cm) was applied. In a subgroup of eight studies,^{14,20-23,26,32,42} with a cutoff of 6.0 cm or less, pooled positive and negative likelihood ratios were updated to 4.1 (95% CI, 2.3-7.0) and 0.8 (95% CI, 0.6-0.9), respectively, with significant heterogeneity, indicating that a thyromental distance of 6.0 cm or less slightly improved the prediction of difficult intubation.

We calculated posttest probability because it enabled us to generalize our results for varying previous incidence.^{10,11} Calculation of posttest probabilities by means of likelihood ratios is shown in table 2. For example, patients with a 5% pretest probability of difficult

intubation have a 15% risk of difficult intubation after a positive thyromental distance test result and a 4% risk of difficult intubation after a negative thyromental distance test result. The risk of difficult intubation after positive and negative test results is shown with a possible range of pretest probabilities (table 2).

Symmetry in the funnel plot was confirmed by significant Kendall correlation coefficients of 0.18 ($P = 0.14$) for the Mallampati test and 0.23 ($P = 0.19$) for thyromental distance, which suggests the absence of publication bias.

Mallampati Classification

The Mallampati score may estimate the size of the tongue relative to the oral cavity^{1,4} and may possibly indicate whether displacement of the tongue by the laryngoscope blade is likely to be easy or difficult. In addition, it assesses whether the mouth can be opened adequately to permit intubation. The Mallampati test assesses not only pharyngeal structure but also head and neck mobility. Recent investigation⁵⁰ has suggested that craniocervical extension relates to mouth opening, and limited head or neck mobility may result in a poor Mallampati score. Despite theoretical arguments for this test, poor pooled sensitivity values and relatively moderate specificity values were obtained in our analysis. Positive and negative likelihood ratios were moderate but unsatisfactory for clinical use. Heterogeneity was present in sensitivity and specificity. Heterogeneity and inadequate diagnostic performance may result in part from inconsistency or uncertainty in performing the tests, *e.g.*, the Mallampati test may have been conducted with or without phonation and/or with different head or tongue positions. Some reports omitted descriptions of how the tests were administered. Because of these factors, the Mallampati test may be of marginal diagnostic value.

Thyromental Distance

Thyromental distance is considered to be an indicator of mandibular space.⁴ This test also reflects whether displacement of the tongue by the laryngoscope blade will be easy or difficult. The diagnostic value of thyromental distance proved unsatisfactory in our analysis. A wide range in test sensitivity may result in heterogeneity. Heterogeneity may be due to the variety of test thresholds: cutoff points varied from 4.0 to 7.0 cm. The summary ROC analysis showed a trend toward variation in overall diagnostic performance of the thyromental test in relation to test threshold. Our additional analysis showed that the positive likelihood ratio improved from 3.4 to 4.1 when a stricter cutoff criterion (≤ 6.0 cm) was applied. Because one study²¹ with a cutoff of less than 4.0 cm yielded higher diagnostic performance with positive and negative likelihood ratios of 9.4 and 0.03, respectively, we should reevaluate the test threshold for

Table 1. Studies Included in the Meta-analysis and Related Data

Study	No. of Patients	Mean Age, yr	Subjects	Diagnostic Criteria for Difficult Intubation	Type of Laryngoscope Blade
Ezri <i>et al.</i> , ¹⁴ 2003	1,472	44.5	Morbidly obese and nonobese	Cormack–Lehane grade III or IV	Macintosh No. 3
Gupta <i>et al.</i> , ¹⁵ 2003	372	16–37	Obstetric	Cormack–Lehane grade III or IV	Macintosh
Ihohom <i>et al.</i> , ¹⁶ 2003	212	>18	General population	Cormack–Lehane grade III or IV, or a patients required bougie in grade 2	Macintosh
Juvin <i>et al.</i> , ¹⁷ 2003	129	40	Obese (BMI \geq 35)	Intubation Difficulty Scale \geq 5	Macintosh No. 3
Khan <i>et al.</i> , ¹⁸ 2003	300	\geq 16	General population	Cormack–Lehane grade III or IV	Macintosh No. 3 or Miller
Brodsky <i>et al.</i> , ¹⁹ 2002	100	44	Morbidly obese (BMI > 40)	Number of intubation attempts \geq 3	Macintosh No. 3 or 4, Miller No. 2
Koh <i>et al.</i> , ²⁰ 2002	605	45	General population	Modified Cormack–Lehane grade \geq IIb, III and IV	Macintosh No. 3
Ayoub <i>et al.</i> , ²¹ 2000	160	NS	General population	Cormack–Lehane grade III or IV	Macintosh
Vani <i>et al.</i> , ²² 2000	50	57.1	Diabetic	Cormack–Lehane grade III or IV	Macintosh
Suyama <i>et al.</i> , ²³ 1999	476	55	General population	Cormack–Lehane grade III or IV	NS
Wong and Hung, ²⁴ 1999	411	38.3	General and pregnant	Cormack–Lehane grade III or IV	NS
Arne <i>et al.</i> , ²⁵ 1998	1,200	>15	Surgery for ENT and general surgery	Unusual techniques required	Macintosh
Bilgin and Ozyurt, ²⁶ 1998	500	46	General	Special instruments required or at least three unsuccessful attempts	Macintosh No. 3
Ulrich <i>et al.</i> , ²⁷ 1998	1,993	49	General	Wilson grades 4 and 5	Macintosh No. 3
Voyagis <i>et al.</i> , ²⁸ 1998	1,833 (99)	NS	Normal and obese (5.4%) Obese only	Cormack–Lehane grade III or IV Cormack–Lehane grade III or IV	Macintosh No. 3 or 4 Macintosh No. 3 or 4
Bergler <i>et al.</i> , ²⁹ 1997	91	NS	Normal	Cormack–Lehane grade III or IV	NS
Yamamoto <i>et al.</i> , ³⁰ 1997	3,680	NS	General	Cormack–Lehane grade III or IV	Macintosh No. 3 or 4
Ramadhani <i>et al.</i> , ³¹ 1996	523	30.4	Obstetric	Cormack–Lehane grade III or IV	Macintosh
El-Ganzouri <i>et al.</i> , ³² 1996	10,507	>18 yr	Normal	Cormack–Lehane grade III or IV	Macintosh and Miller
Jacobsen <i>et al.</i> , ³³ 1996	100	NS	Normal	No epiglottis or only soft palate visible	Macintosh
Laplace <i>et al.</i> , ³⁴ 1995	849		General	Special procedure required (external compression excepted)	NS
Samra <i>et al.</i> , ³⁵ 1995	564	NS	Normal	Cormack–Lehane grade III or IV	NS
Tse <i>et al.</i> , ³⁶ 1995	471	\geq 18 yr	Normal	Cormack–Lehane grade III or IV	Macintosh No. 3
Descoins <i>et al.</i> , ³⁷ 1994	295	NS	Normal and tumor	Cormack–Lehane grade III or IV, or special instruments required	Macintosh No. 3 or Miller
Ita <i>et al.</i> , ³⁸ 1994	57	30.5	Normal	Cormack–Lehane grade III or IV	Macintosh No. 3
Rose and Cohen, ³⁹ 1994	18,205	NS	Normal (obstetric excluded)	Recorded as difficult with more than two laryngoscopies required)	NS
Sava <i>et al.</i> , ⁴⁰ 1994	350	58	Obstetric and nonobstetric	Cormack–Lehane grade III or IV, or use of a gum elastic bougie	Macintosh
Restelli <i>et al.</i> , ⁴¹ 1993	700	NS	Normal	Cormack–Lehane grade III or IV	NS
Butler and Dhara, ⁴² 1992	250	16–80	Normal	Cormack–Lehane grade III or IV	Macintosh No. 3
Cohen <i>et al.</i> , ⁴³ 1992	663	18–88	Normal	Cormack–Lehane grade III or IV	NS
Rocke <i>et al.</i> , ⁴⁴ 1992	1,500	26.4	Obstetric	Cormack–Lehane grade III or IV	NS
Yeo <i>et al.</i> , ⁴⁵ 1992	283	NS	Gynecologic	Cormack–Lehane grade III or IV	NS
	277		Obstetric	Cormack–Lehane grade III or IV	NS
Frerk, ⁴⁶ 1991	244	44.3	Normal	Cormack–Lehane grade III or IV, or a gum elastic bougie required	Macintosh
Oates <i>et al.</i> , ⁴⁷ 1991	675	46.6	Normal	Only epiglottis visible or not even epiglottis visible (Wilson)	NS
Pottecher <i>et al.</i> , ⁴⁸ 1991	663	NS	Normal	Cormack–Lehane grade III or IV	Macintosh

BMI = body mass index; CI = confidence interval; ENT = ear, nose, and throat; MP = Mallampati classification; NS = not specified; OPC = oropharyngeal class; SMD = sternal distance; TMD = thyromental distance.

Table 1. Continued

No. of Patients with Difficult Intubation	Incidence of Difficult Intubation, %	Predictor	Sensitivity (95% CI)	Specificity (95% CI)	Quality Score
152	10.3	MP grades III and IV	0.76 (0.69–0.82)	0.73 (0.71–0.75)	3
		TMD <6.0 cm	0.30 (0.24–0.38)	0.86 (0.84–0.88)	
25	6.7	MP grades III and IV	0.60 (0.41–0.76)	0.98 (0.95–0.99)	3
		Wilson risk score ≥ 2	0.37 (0.21–0.56)	0.98 (0.96–0.99)	
20	9.0	MP grades III and IV	0.40 (0.23–0.61)	0.88 (0.83–0.92)	2
		TMD <6.5 cm	0.45 (0.26–0.66)	0.95 (0.90–0.97)	
		SMD <12.5 cm	0.40 (0.23–0.61)	0.97 (0.94–0.99)	
		Combination (MP grade III and IV + TMD <6.5 cm)	0.26 (0.12–0.48)	1.00 (0.98–1.00)	
20	15.5	MP grades III and IV	0.83 (0.63–0.94)	0.62 (0.53–0.71)	3
		Mouth opening <35 mm	0.40 (0.23–0.61)	0.76 (0.67–0.83)	
17	5.7	MP grades III and IV	0.81 (0.58–0.93)	0.67 (0.61–0.72)	4
12	12.0	MP grades III and IV	0.58 (0.32–0.80)	0.70 (0.60–0.79)	3
31	5.1	MP grades III and IV	0.45 (0.30–0.62)	0.92 (0.90–0.94)	1
		TMD <6.0 cm	0.42 (0.27–0.59)	0.88 (0.85–0.90)	
21	13.2	TMD <4.0 cm	0.98 (0.82–1.00)	0.90 (0.83–0.94)	2
		Combination (MP + TMD)	0.52 (0.33–0.71)	0.98 (0.93–0.99)	
8	16.0	TMD <6.0 cm	0.28 (0.09–0.60)	0.94 (0.83–0.98)	2
26	5.5	MP grades II, III, and IV	0.65 (0.46–0.80)	0.75 (0.71–0.79)	3
		TMD ≤ 6.0 cm	0.57 (0.39–0.74)	0.63 (0.58–0.67)	
7	1.7	MP grades III and IV	0.81 (0.47–0.96)	0.63 (0.58–0.67)	3
		TMD <6.5 cm	0.94 (0.60–0.99)	0.24 (0.20–0.29)	
		Combination (MP + TMD)	0.81 (0.47–0.96)	0.67 (0.62–0.71)	
50	4.2	TMD <6.5 cm	0.17 (0.09–0.29)	0.95 (0.93–0.96)	2
		MP grades III and IV	0.77 (0.64–0.87)	0.85 (0.83–0.87)	
40	8.0	MP grade III	0.43 (0.29–0.58)	0.93 (0.90–0.95)	1
		TMD <6.0 cm	0.35 (0.23–0.51)	0.95 (0.93–0.97)	
		Wilson risk score > 2	0.57 (0.42–0.71)	0.91 (0.88–0.93)	
94	4.7	MP grade III	0.17 (0.11–0.26)	0.93 (0.92–0.94)	2
		TMD <7.0	0.16 (0.10–0.25)	0.86 (0.84–0.87)	
152	8.3	MP grades III and IV	0.87 (0.80–0.91)	0.92 (0.91–0.93)	2
(20)	20.2	MP grades III and IV	0.88 (0.68–0.96)	0.88 (0.79–0.94)	
10	11.0	MP grades III and IV	0.59 (0.32–0.82)	0.71 (0.61–0.80)	3
56	1.5	MP grades III and IV	0.68 (0.55–0.78)	0.52 (0.51–0.54)	4
		Wilson risk score ≥ 2	0.55 (0.42–0.67)	0.86 (0.85–0.87)	
18	3.4	SMD ≤ 13.5 cm	0.66 (0.43–0.83)	0.71 (0.67–0.75)	3
642	6.1	MP grades III and IV	0.45 (0.41–0.49)	0.89 (0.88–0.90)	3
		TMD <6.0 cm	0.07 (0.05–0.09)	0.99 (0.99–0.99)	
		Mouth opening	0.26 (0.23–0.30)	0.95 (0.94–0.95)	
7	7.0	MP grades III and IV	0.94 (0.60–0.99)	0.84 (0.75–0.90)	2
62	7.3	MP grades III and IV	0.34 (0.24–0.46)	0.89 (0.86–0.91)	2
		Wilson risk score ≥ 2	0.36 (0.25–0.48)	0.92 (0.89–0.93)	
48	8.5	MP grades III and IV	0.50 (0.37–0.63)	0.87 (0.83–0.89)	2
62	13.2	MP grades III and IV	0.66 (0.54–0.76)	0.65 (0.60–0.69)	4
		TMD <7.0 cm	0.33 (0.22–0.45)	0.80 (0.76–0.83)	
		Combination (MP + TMD)	0.21 (0.13–0.33)	0.92 (0.89–0.94)	
41	14.0	TMD <6.5 cm	0.25 (0.14–0.40)	0.92 (0.88–0.95)	2
3	5.3	MP grades III and IV	0.63 (0.22–0.91)	0.95 (0.86–0.99)	2
326	1.8	Mouth opening <2 fingers	0.05 (0.03–0.09)	0.99 (0.99–0.99)	2
		TMD <3 fingers	0.15 (0.11–0.21)	0.98 (0.97–0.98)	
		Uvula not seen (= MP grade IV)	0.11 (0.08–0.17)	0.96 (0.96–0.97)	
17	4.9	MP grades III and IV	0.64 (0.41–0.82)	0.67 (0.61–0.71)	3
		TMD ≤ 6.5 cm	0.64 (0.41–0.82)	0.79 (0.74–0.83)	
		SMD ≤ 12.5 cm	0.81 (0.58–0.93)	0.88 (0.85–0.91)	
77	11.0	MP grades III and IV	0.33 (0.23–0.44)	0.92 (0.90–0.94)	2
18	7.2	MP grades III and IV	0.55 (0.34–0.75)	0.81 (0.75–0.86)	2
		TMD <6.0 cm	0.61 (0.39–0.79)	0.75 (0.69–0.81)	
76	11.5	MP grades III and IV	0.28 (0.19–0.39)	0.95 (0.93–0.96)	1
27	1.8	MP grades III and IV	0.59 (0.41–0.75)	0.74 (0.72–0.76)	1
5	1.8	MP grades III and IV	0.36 (0.12–0.70)	0.96 (0.93–0.98)	3
6	2.2	MP grades III and IV	0.25 (0.06–0.64)	0.95 (0.92–0.97)	
11	4.5	MP grades III and IV	0.79 (0.51–0.93)	0.81 (0.76–0.86)	1
		TMD <7.0 cm	0.88 (0.60–0.97)	0.81 (0.76–0.86)	
		Combined (MP + TMD)	0.79 (0.51–0.93)	0.98 (0.95–0.99)	
12	1.8	MP grades III and IV	0.42 (0.20–0.68)	0.84 (0.80–0.86)	2
		Wilson risk score ≥ 2	0.42 (0.20–0.68)	0.92 (0.90–0.94)	
39	5.9	MP grades III and IV	0.44 (0.30–0.59)	0.97 (0.96–0.98)	3

Table 2. Pooled Estimates of Bayesian Statistics of Six Different Bedside Tests for Difficult Intubation

Diagnostic Test	No. of Studies Included	No. of Patients	Prevalence of Difficult Intubation (95% CI), %	Pooled Sensitivity (95% CI), %	Pooled Specificity (95% CI), %	Pooled Likelihood Ratio (95% CI)		Pooled Log Diagnostic Odds Ratio (95% CI)
						Pos.	Neg.	
Overall population								
Mallampati classification	31	41,193	5.7 (4.4–7.3)*	49 (41–57)*	86 (81–90)*	3.7 (3.0–4.6)*	0.5 (0.5–0.6)*	2.0 (1.7–2.3)*
Thyromental distance	17	29,132	6.5 (4.6–9.1)*	20 (11–29)*	94 (89–99)*	3.4 (2.3–4.9)*	0.8 (0.8–0.9)*	1.7 (1.2–2.1)*
Sternomental distance	3	1,085	5.4 (3.1–9.2)*	62 (37–86)*	82 (67–97)*	5.7 (2.1–15.1)*	0.5 (0.3–0.8)	2.7 (1.4–3.9)*
Mouth opening	3	20,614	5.6 (2.2–14.5)*	22 (9–35)*	97 (93–100)*	4.0 (2.0–8.2)*	0.8 (0.7–1.0)*	1.7 (1.2–2.3)*
Wilson risk score	5	6,076	4.0 (1.8–9.0)*	46 (36–56)	89 (85–92)	5.8 (3.9–8.6)*	0.6 (0.5–0.9)	2.3 (1.8–2.8)*
Combination of Mallampati classification and thyromental distance	5	1,498	6.6 (2.8–15.6)*	36 (14–59)*	87 (74–100)*	9.9 (3.1–31.9)*	0.6 (0.5–0.9)*	3.3 (1.5–5.0)*
Obstetric subgroup								
Mallampati classification	3	2,155	3.1 (1.7–5.5)*	56 (41–72)	81 (67–95)*	6.4 (1.1–36.5)*	0.6 (0.4–0.8)	2.5 (0.6–4.4)*
Obese subgroup (BMI > 30)								
Mallampati classification	4	378	15.8 (14.3–17.5)	74 (51–97)*	74 (62–87)*	2.9 (1.6–5.3)*	0.4 (0.2–0.8)	2.1 (0.8–3.3)*

Posttest probability = [(pretest odds) * likelihood ratio]/[1 + (pretest odds) * likelihood ratio]; where pretest odds = pretest probability/(1 – pretest probability).

DerSimonian–Laird random-effects model was used throughout. * Significant heterogeneity ($P < 0.1$) was found.

BMI = body mass index; CI = confidence interval; Neg. = negative; Pos. = positive; ROC = receiver operating characteristic curve.

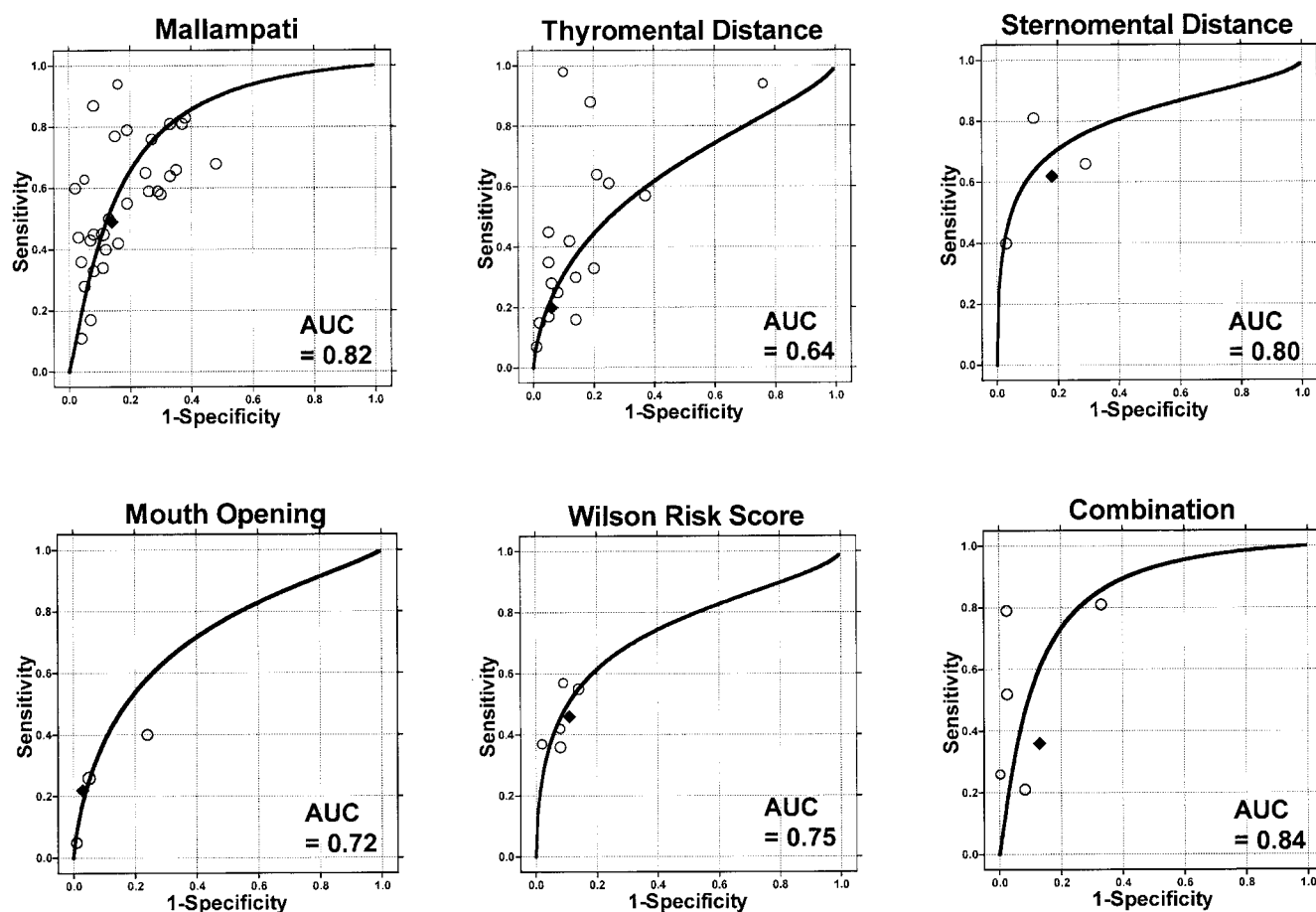


Fig. 2. Summary receiver operating characteristic curve analysis of six different bedside screening tests for difficult intubation. Weighted summary receiver operating characteristic curve is expressed by a solid line. Individual study estimates of sensitivity and (1 – specificity) are shown by open circles. Each circle is proportional to the inverse of the variance. Diamonds indicate pooled point estimates of sensitivity and specificity. AUC = area under the curve; Combination = Mallampati classification and thyromental distance in combination.

Table 2. Continued

Best Regression Model for Summary ROC ($D = a + bS$)			Posttest Probability, %					
			Pretest Probability = 5%		Pretest Probability = 10%		Pretest Probability = 15%	
a	b	P	Pos. Result	Neg. Result	Pos. Result	Neg. Result	Pos. Result	Neg. Result
2.1	0.08	0.39	16	3	29	6	39	9
0.9	-0.16	0.06	15	4	27	8	37	13
2.1	-0.29	0.65	23	2	39	5	50	8
1.4	-0.11	0.59	18	4	31	9	42	13
1.6	-0.24	0.48	23	3	39	6	50	10
2.4	0.07	0.92	34	3	52	6	64	10
1.0	-0.70	0.42	25	3	42	6	53	9
2.1	0.41	0.55	13	2	24	4	34	7

thyromental distance. Another source of heterogeneity may be variation in measurement conditions: Thyromental distance could have been measured from inside or outside the mentum. The methods of measurement must be standardized.

Sternomental Distance

Sternomental distance can be an indicator of head and neck mobility.³¹ Head extension is believed to be an important factor in determining the ease or difficulty of intubation. Among single-factor tests, sternomental distance yielded the highest positive likelihood ratio and diagnostic odds ratio with moderate sensitivity and specificity. The negative likelihood ratio was lower than that of any other test, suggesting that it is the best single test for ruling out difficult intubation. The cutoff point of sternomental distance was consistently 12.5 to 13.5 cm. However, only three studies were included in our analysis. Therefore, the diagnostic performance remains inconclusive. Further investigation is required because so few studies address sternomental distance.

Mouth Opening

Mouth opening seemed in our analysis to be an inadequate predictor of difficult intubation. It may be argued that mouth opening indicates movement of the temporomandibular joint and that significantly limited mouth opening hinders exposure of the larynx. Several studies based on multivariate analysis^{3,51} indicated that limited mouth opening is strongly associated with difficult intubation. Unexpected results may have been obtained in our analysis, because measurement thresholds varied. The threshold was even unclear in one study.³² Our analysis suggests that mouth opening is not a useful test; however, we could not determine whether this is because of limited data or because mouth opening is truly

not useful in predicting difficult intubation. This area would benefit from further investigation.

Wilson Risk Score

The CI of the Wilson risk score is narrower than that of other tests, and sensitivity and specificity are homogeneous. The same criterion (score ≥ 2) was applied in all studies included in our analysis, making the data cluster very closely together and thus yielding a narrower CI in pooled sensitivity and specificity. All included studies set the test threshold somewhat high; therefore, sensitivity remained low and specificity remained high on our summary ROC curve. The Wilson risk score with a cutoff value of 2 or greater yielded a low true-positive rate and a low false-positive rate, meaning that the test threshold correctly identifies patients for whom intubation will be easy. Although our analysis did not include the original data of Wilson *et al.*,³ our pooled sensitivity and specificity with a cutoff score of 2 or greater seem to be similar to their original sensitivity and specificity data. This suggests that the Wilson risk score has high reproducibility.

Combination of Mallampati Classification and Thyromental Distance

We found that a combination of the Mallampati test and thyromental distance most accurately predicted difficult intubation. This combination yielded low sensitivity, but the positive likelihood ratio (9.9) supports the test as a strong predictor of difficult intubation. The diagnostic odds ratio (3.3) and the area under the summary ROC curve (0.84) are the highest of all tests. Patients with a 5% pretest probability of difficult intubation were shown to have a 34% risk of difficult intubation after a positive result for the combination test, a 16% risk after a positive result of Mallampati test alone, and a 15%

risk after a positive result of thyromental distance alone. Therefore, the discriminative power is greater when the tests are used in combination rather than alone. It is suggested that a combination of the Mallampati classification and thyromental distance has the highest discriminative power among currently available tests. However, heterogeneity and an insufficient number of studies limit definitive conclusions.

Mallampati Classification in Obstetric and Obese Populations

We found that diagnostic performance of the Mallampati test in obstetric and obese populations is similar to that in the overall population. The diagnostic odds ratios in these populations are similar, and the trend toward poor sensitivity and fair specificity remained. We also found the incidence of difficult intubation in obese (body mass index > 30) patients to be more than three times that of normal patients. Obese patients with a 15% pretest probability of difficult intubation had a 34% risk of difficult intubation after a positive Mallampati test result, twice the risk of the normal population with a 5% pretest probability. Excessive soft tissue in the velopalate, retropharynx, and submandibular regions in obese patients may cause difficulty in laryngoscopy.⁴⁹ Our result confirms the common understanding that obese patients have a greater incidence of difficult intubation than that of normal patients. Because of the high incidence of difficult intubation in these patients, the Mallampati test may yield higher posttest probability of difficult intubation in obese patients than in normal patients. Data for obstetric population, however, remain inconclusive because of the small number of studies and the heterogeneity.

Strengths and Limitations

Our meta-analysis showed the incidence of difficult intubation in normal patients without pathologic airway anatomy to be 5.8%, which lies within the limits of the incidence reported in the literature we reviewed.^{2,4,52,53} This can be viewed as a strength in terms of the external validity of our findings. However, our meta-analysis has several limitations. First, publication bias was not identified for the Mallampati classification and for thyromental distance. However, few studies were included for the other diagnostic tests; there may be unpublished studies. Second, the reference standard for difficult intubation differed somewhat among studies. Most studies defined difficult intubation as a Cormack–Lehane grade of 3 or greater, but some studies used other classification systems (e.g., Intubation Difficulty Scale Score¹⁷) or repeated attempts.^{19,26,40} The Cormack–Lehane scale was not originally designed for grading the degree of difficulty in laryngoscopy or tracheal intubation.⁵ In addition, laryngoscopy with or without application of external cricoid pressure or of backward, upward, and

rightward pressure (BURP maneuver) on the thyroid cartilage to facilitate a laryngoscopic view might have affected the Cormack–Lehane grade in individual studies. Controversy lingers as to the definitions of difficult intubation and difficult laryngoscopy.^{5,53}

Given that screening tests included proved to have inadequate diagnostic power, is any attempt at prediction likely to be useful? Should any predictive attempt be advocated? This question cannot be generally answered; however, as Wilson stated, “No test is likely to be perfect, therefore, it remains essential that every anesthetist must be trained and equipped to deal with the now much less common, unexpected failure to intubate.”⁵⁴ We concur, and we believe that attempts at prediction are much less important than knowing what to do when difficulty is encountered.

In conclusion, currently available screening tests for difficult intubation have only poor to moderate discriminative power when used alone. Combinations of individual tests or risk factors add some incremental diagnostic value in comparison to the value of each test alone. However, the clinical value of these bedside screening tests for predicting difficult intubation remains limited.

The authors thank Joseph Lau, M.D. (Professor of Medicine and Clinical Research and Director of the Center for Clinical Evidence Synthesis, Tufts-New England Medical Center, Boston, Massachusetts), for providing statistical advice and reviewing the manuscript and Javier Zamora, M.D. (Associate Professor, Unidad de Bioestadística Clínica, Hospital Ramón y Cajal, Madrid, Spain), for his courtesy of providing statistical software. We also thank Toshiro Shitara, M.D., Ph.D. (Chief Anesthesiologist, Sakakibara Memorial Hospital, Tokyo, Japan), for his superb technical assistance.

References

- Mallampati SR, Gatt SP, Gugino LD, Desai SP, Waraksa B, Freiburger D, Liu PL: A clinical sign to predict difficult tracheal intubation: A prospective study. *Can Anaesth Soc J* 1985; 32:429–34
- Janssens M, Hartstein G: Management of difficult intubation. *Eur J Anaesthesiol* 2001; 18:3–12
- Wilson ME, Spiegelhalter D, Robertson JA, Lesser P: Predicting difficult intubation. *Br J Anaesth* 1988; 61:211–6
- Randell T: Prediction of difficult intubation. *Acta Anaesthesiol Scand* 1996; 40:1016–23
- Yentis SM: Predicting difficult intubation: Worthwhile exercise or pointless ritual? *Anaesthesia* 2002; 57:105–9
- Romagnuolo J, Bardou M, Rahme E, Joseph L, Reinhold C, Barkun AN: Magnetic resonance cholangiopancreatography: A meta-analysis of test performance in suspected biliary disease. *Ann Intern Med* 2003; 139:547–57
- Cormack RS, Lehane J: Difficult tracheal intubation in obstetrics. *Anaesthesia* 1984; 39:1105–11
- DerSimonian R, Laird N: Meta-analysis in clinical trials. *Control Clin Trials* 1986; 7:177–88
- Altman DG, Bland JM: Diagnostic tests: 1. Sensitivity and specificity. *BMJ* 1994; 308:1552
- Deeks JJ, Altman DG: Diagnostic tests: 4. Likelihood ratios. *BMJ* 2004; 329:168–9
- Deeks JJ: Systematic reviews of evaluations of diagnostic and screening tests, *Systematic Reviews in Health Care: Meta-Analysis in Context*, 2nd edition. Edited by Egger M, Smith GD, Altman DG. London, BMJ Books, 2001, pp 248–82
- Moses LE, Shapiro D, Littenberg B: Combining independent studies of a diagnostic test into a summary ROC curve: Data-analytic approaches and some additional considerations. *Stat Med* 1993; 12:1293–316
- Begg CB, Mazumdar M: Operating characteristics of a rank correlation test for publication bias. *Biometrics* 1994; 50:1088–101
- Ezri T, Medalion B, Weisenberg M, Szmuk P, Warters RD, Charuzi I:

- Increased body mass index per se is not a predictor of difficult laryngoscopy. *Can J Anaesth* 2003; 50:179-83
15. Gupta S, Pareek S, Dulara SC: Comparison of two methods for predicting difficult intubation in obstetric patients. *Middle East J Anesthesiol* 2003; 17: 275-85
 16. Iohom G, Ronayne M, Cunningham AJ: Prediction of difficult tracheal intubation. *Eur J Anaesthesiol* 2003; 20:31-6
 17. Juvin P, Lavaut E, Dupont H, Lefevre P, Demetriou M, Dumoulin JL, Desmonts JM: Difficult tracheal intubation is more common in obese than in lean patients. *Anesth Analg* 2003; 97:595-600
 18. Khan ZH, Kashfi A, Ebrahimkhani E: A comparison of the upper lip bite test (a simple new technique) with modified Mallampati classification in predicting difficulty in endotracheal intubation: A prospective blinded study. *Anesth Analg* 2003; 96:595-9
 19. Brodsky JB, Lemmens HJ, Brock-Utne JG, Vierra M, Saidman LJ: Morbid obesity and tracheal intubation. *Anesth Analg* 2002; 94:732-6
 20. Koh LK, Kong CE, Ip-Yam PC: The modified Cormack-Lehane score for the grading of direct laryngoscopy: Evaluation in the Asian population. *Anaesth Intensive Care* 2002; 30:48-51
 21. Ayoub C, Baraka A, el-Khatib, M, Muallem M, Kawkabani N, Soueida A: A new cut-off point of thyromental distance for prediction of difficult airway. *Middle East J Anesthesiol* 2000; 15:619-33
 22. Vani V, Kamath SK, Naik LD: The palm print as a sensitive predictor of difficult laryngoscopy in diabetics: A comparison with other airway evaluation indices. *J Postgrad Med* 2000; 46:75-9
 23. Suyama H, Tsuno S, Takeyoshi S: The clinical usefulness of predicting difficult endotracheal intubation [in Japanese]. *Masui* 1999; 48:37-41
 24. Wong SH, Hung CT: Prevalence and prediction of difficult intubation in Chinese women. *Anaesth Intensive Care* 1999; 27:49-52
 25. Arne J, Descoins P, Fusciardi J, Ingrand P, Ferrier B, Boudigues D, Aries 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
 26. Bilgin H, Ozyurt G: Screening tests for predicting difficult intubation: A clinical assessment in Turkish patients. *Anaesth Intensive Care* 1998; 26:382-6
 27. Ulrich B, Listyo R, Gerig HJ, Gabi K, Kreienbuhl G: The difficult intubation: The value of BURP and 3 predictive tests of difficult intubation [in German]. *Anaesthesist* 1998; 47:45-50
 28. Voyagis GS, Kyriakis KP, Dimitriou V, Vrettou I: Value of oropharyngeal Mallampati classification in predicting difficult laryngoscopy among obese patients. *Eur J Anaesthesiol* 1998; 15:330-4
 29. Bergler W, Maleck W, Baker-Schreyer A, Ungemach J, Petroianu G, Hornmann K: The Mallampati score: Prediction of difficult intubation in otolaryngologic laser surgery by Mallampati Score [in German]. *Anaesthesist* 1997; 46: 437-40
 30. Yamamoto K, Tsubokawa T, Shibata K, Ohmura S, Nitta S, Kobayashi T: Predicting difficult intubation with indirect laryngoscopy. *ANESTHESIOLOGY* 1997; 86:316-21
 31. Al Ramadhani S, Mohamed LA, Rocke DA, Gouws E, Ramadhani SA: Sternomental distance as the sole predictor of difficult laryngoscopy in obstetric anaesthesia. *Br J Anaesth* 1996; 77:312-6
 32. 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
 33. Jacobsen J, Jensen E, Waldau T, Poulsen TD: Preoperative evaluation of intubation conditions in patients scheduled for elective surgery. *Acta Anaesthesiol Scand* 1996; 40:421-4
 34. Laplace E, Benefice S, Marti Flich J, Patrigeon RG, Combourieu E: Difficult intubation: A prospective evaluation of the Mallampati and Wilson tests [in French]. *Cah Anesthesiol* 1995; 43:205-8
 35. Samra SK, Schork MA, Guinto Jr FC: A study of radiologic imaging techniques and airway grading to predict a difficult endotracheal intubation. *J Clin Anesth* 1995; 7:373-9
 36. Tse JC, Rimm EB, Hussain A: Predicting difficult endotracheal intubation in surgical patients scheduled for general anesthesia: A prospective blind study. *Anesth Analg* 1995; 81:254-8
 37. Descoins P, Arne J, Bresard D, Aries J, Fusciardi J: Proposal for a new multifactor screening score of difficult intubation in ORL and stomatognathic surgery: Preliminary study [in French]. *Ann Fr Anesth Reanim* 1994; 13:195-200
 38. Ita CE, Eshiet AI, Akpan SG: Recognition of the difficult airway in normal Nigerian adults (a prospective study). *West Afr J Med* 1994; 13:102-4
 39. Rose DK, Cohen MM: The airway: Problems and predictions in 18,500 patients. *Can J Anaesth* 1994; 41:372-83
 40. Savva D: Prediction of difficult tracheal intubation. *Br J Anaesth* 1994; 73:149-53
 41. Restelli L, Moretti MP, Todaro C, Banfi L: The Mallampati's scale: A study of reliability in clinical practice [in Italian]. *Minerva Anesthesiol* 1993; 59:261-5
 42. Butler PJ, Dhara SS: Prediction of difficult laryngoscopy: An assessment of the thyromental distance and Mallampati predictive tests. *Anaesth Intensive Care* 1992; 30:31-2
 43. 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* 1992; 4:310-4
 44. Rocke DA, Murray WB, Rout CC, Gouws E: Relative risk analysis of factors associated with difficult intubation in obstetric anaesthesia. *ANESTHESIOLOGY* 1992; 77:67-73
 45. Yeo SW, Chong JL, Thomas E: Difficult intubation: A prospective study. *Singapore Med J* 1992; 33:362-4
 46. Frerk CM: Predicting difficult intubation. *Anaesthesia* 1991; 46:1005-8
 47. Oates JD, Macleod AD, Oates PD, Pearsall FJ, Howie JC, Murray GD: Comparison of two methods for predicting difficult intubation. *Br J Anaesth* 1991; 66:305-9
 48. Pottecher T, Velten M, Galani M, Forriker M: Comparative value of clinical signs of difficult tracheal intubation in women [in French]. *Ann Fr Anesth Reanim* 1991; 10:430-5
 49. Ezri T, Gewurtz G, Sessler DI, Medalion B, Szmuk P, Hagberg C, Susmallian S: Prediction of difficult laryngoscopy in obese patients by ultrasound quantification of anterior neck soft tissue. *Anaesthesia* 2003; 58:1111-4
 50. Calder I, Picard J, Chapman M, O'Sullivan, C, Crockard HA: Mouth opening: A new angle. *ANESTHESIOLOGY* 2003; 99:799-801
 51. Karkouti K, Rose DK, Wigglesworth D, Cohen MM: Predicting difficult intubation: A multivariable analysis. *Can J Anaesth* 2000; 47:730-9
 52. Practice Guidelines for Management of the Difficult Airway: An updated report by the American Society of Anesthesiologists Task Force on Management of the Difficult Airway. *ANESTHESIOLOGY* 2003; 98:1269-77
 53. Crosby ET, Cooper RM, Douglas MJ, Doyle DJ, Hung OR, Labrecque P, Muir H, Murphy MF, Preston RP, Rose DK, Roy L: The unanticipated difficult airway with recommendations for management. *Can J Anaesth* 1998; 45:757-76
 54. Wilson ME: Predicting difficult intubation. *Br J Anaesth* 1993; 71:333-4