A Systematic Review (Meta-Analysis) of the Accuracy of the Mallampati Tests to Predict the Difficult Airway

Anna Lee, PhD, MPH, Lawrence T. Y. Fan, MBBS, FANZCA, Tony Gin, MD, FRCA, FANZCA, MB, ChB, Manoj K. Karmakar, MD, FRCA, and Warwick D. Ngan Kee, MD, FANZCA, MB, ChB

Department of Anaesthesia and Intensive Care, The Chinese University of Hong Kong, Prince of Wales Hospital, Shatin, NT, Hong Kong

The original and modified Mallampati tests are commonly used to predict the difficult airway, but there is controversy regarding their accuracy. We searched MEDLINE and other databases for prospective studies of patients undergoing general anesthesia in which the results of a preoperative Mallampati test were compared with the subsequent rate of difficult airway (difficult laryngoscopy, difficult intubation, or difficult ventilation as reference tests). Forty-two studies enrolling 34,513 patients were included. The definitions of the reference tests varied widely. For predicting difficult laryngoscopy, both versions of the Mallampati test had

ifficult laryngoscopy and difficult tracheal intubation occur in 1.5% to 8% of general anesthetics (1). Of available methods, the Mallampati, original (2) and modified (3,4) tests, are used as a preoperative bedside test to predict a difficult airway (5). However, the usefulness of this test is unclear, as published studies have produced variable estimates of diagnostic test accuracy. The original Mallampati test (2) identified difficult intubations with a high degree of accuracy, with sensitivity of 50% and specificity of 100%. However, subsequent larger studies have shown only modest degrees of accuracy using the original (6) and modified (7-9) versions of the test. Furthermore, the accuracy of the Mallampati test may vary according to patients' ethnic group and sex and whether they are pregnant (10). For example, in Asian patients it may be more difficult to intubate the trachea than in Caucasians (11,12).

@2006 by the International Anesthesia Research Society 0003-2999/06

good accuracy (area under the summary receiver operating characteristic (sROC) curve = 0.89 ± 0.05 and 0.78 ± 0.05 , respectively). For predicting difficult intubation, the modified Mallampati test had good accuracy (area under the sROC curve = 0.83 ± 0.03) whereas the original Mallampati test was poor (area under the sROC curve = 0.58 ± 0.12). The Mallampati tests were poor at identifying difficult mask ventilation. Publication bias was not detected. Used alone, the Mallampati tests have limited accuracy for predicting the difficult airway and thus are not useful screening tests.

(Anesth Analg 2006;102:1867–78)

The objective of this systematic review was to determine the accuracy of the Mallampati test for predicting the difficult airway. For the purposes of this review, the definition of a difficult airway included difficult laryngoscopy, difficult tracheal intubation, and difficult ventilation. The null hypothesis tested was that all versions of the Mallampati test had poor accuracy for identifying difficult airway. We also explored sources of heterogeneity to increase the clinical relevance of the results.

Methods

This systematic review and meta-analysis followed guidelines on conducting systematic reviews of diagnostic studies (13,14). We included all prospective observational studies of patients undergoing general anesthesia who had preoperative Mallampati test assessments and a subsequent assessment of difficult laryngoscopy, difficult tracheal intubation, or difficult mask ventilation. Difficult airway was defined by a grade III score in the original Mallampati test (2) or a grade III or IV in the modified Mallampati test (3,4) (Table 1).

The studies we assessed included patients with no known risk factors for difficult tracheal intubation as

Accepted for publication February 6, 2006.

Supported solely from departmental and institutional funding. Address correspondence to Anna Lee, PhD, MPH, Department of Anaesthesia and Intensive Care, The Chinese University of Hong Kong, Prince of Wales Hospital, Shatin, NT, Hong Kong. Address e-mail to annalee@cuhk.edu.hk.

DOI: 10.1213/01.ane.0000217211.12232.55

*								
Original Mallampati Test (2)	Grade 1	Faucial pillars, soft palate, and uvula could be visualized						
	Grade 2	Faucial pillars and soft palate could be visualized, but uvula was masked by base of the tongue						
	Grade 3	Only soft palate could be visualized						
Modified Mallampati Test (3)	Class 1	Soft palate, fauces, uvula, pillars seen						
	Class 2	Soft palate, fauces, uvula seen						
	Class 3	Soft palate, base of uvula seen						
	Class 4	Soft palate not visible at all						
Modified Mallampati Test (4)	Class 0	Ability to see any part of the epiglottis on mouth opening and tongue protrusion						
	Class 1	Soft palate, fauces, uvula, pillars seen						
	Class 2	Soft palate, fauces, uvula seen						
	Class 3	Soft palate, base of uvula seen						
	Class 4	Soft palate not visible at all						
Cormack and Lehane scale (20)	Grade 1	Full view of the glottis						
	Grade 2	Partial view of the glottis or arytenoids						
	Grade 3	Only epiglottis visible						
	Grade 4	Neither glottis nor epiglottis visible						
Modified Cormack and Lehane scale (21)	Grade 1	Full view of the glottis						
	Grade 2a	Partial view of the glottis						
	Grade 2b	Arytenoids or posterior part of the vocal cords only just visible						
	Grade 3	Only epiglottis visible						
	Grade 4	Neither glottis nor epiglottis visible						

Table 1. Definition of Mallampati Tests and Cormack and Lehane Scales

well as patients with upper airway pathology, diabetes, obesity, and patients who were pregnant. All types of surgery were considered. Patients undergoing indirect laryngoscopy were excluded (8,15). Retrospective studies (3,16) and case-control studies (17– 19) were excluded because these would overestimate the diagnostic test accuracy compared with studies using a prospective clinical population. The reference tests for difficult airway with which the Mallampati test were compared included difficult laryngoscopy (as defined by the four-grade Cormack and Lehane scoring system (20) or the modified five-grade Cormack and Lehane classification (21) and difficult intubation and difficult ventilation (as defined by the authors).

Search Strategy

A systematic search of all relevant prospective observational studies was conducted. Relevant studies were identified from electronic databases (MEDLINE, EMBASE, Science Citation Index, The Cochrane Library) January 1985–December 2004, and reference lists of relevant studies and reviews in major journals related to anesthesia. Articles were restricted to those published in English, as there is no evidence to suggest a strong association between language restriction and publication bias in systematic reviews of diagnostic tests (22). We used four databases to ensure that relevant articles were identified, as publication bias is more likely to be found if only one to two databases are used in systematic reviews of diagnostic tests (22). In addition, the following subject headings and text words, and their combinations, were included in electronic database search strategy: *sensitivity*, *specificity*, *screening*, *false positive*, *false negative*, *predictive value of tests*, *reference values*, *roc analyses*, *roc area*, *roc characteristics*, *roc curve*, *endotracheal intubation*, *intratracheal intubation*, *laryngoscopy*, *difficult laryngoscopy*, *difficult intubation*, *Mallampati* and *Cormack and Lehane*.

The methodological quality of eligible studies was assessed independently under open conditions. Methods of recruitment and blinding between test and reference test results among anesthesiologists were recorded. The patient population, type of surgery and details of test and reference tests were also collected. Data were obtained from studies independently by two or more investigators using a standardized data extraction form; disagreements were resolved by consensus. The primary author was contacted by letter or email for relevant data that were not presented in the original publication.

Outcome Measures

The primary outcomes were 1) difficult laryngoscopy (20,21) (Cormack and Lehane Grades 2b, 3 and 4) (Table 1) and 2) difficult tracheal intubation (as there is no standard definition for difficult intubation, we accepted the definition used by authors from each study). The secondary outcome was difficult ventilation, as defined by authors from each study. The primary and secondary outcomes were chosen because they are related to consensus guidelines (23) and are clinically important measures of difficult airway.

Statistical Analysis

The sensitivity, specificity, positive likelihood ratio, and negative likelihood ratio were determined individually from each included study. The accuracy of the test was judged by the magnitude of the positive and negative likelihood ratios (how much a given diagnostic test result will increase or decrease the pre-test probability of the target disorder) using the guide by Jaeschke et al. (24). The potential problems associated with sensitivities and specificities of 0% and 100% were solved by adding 0.5 to all cells of the diagnostic 2×2 table (13). The DerSimonian Laird method (random effects model) was used to incorporate variation among studies when pooling sensitivity, specificity, positive likelihood ratio, and negative likelihood ratio. However, when there was an association between sensitivity and specificity across studies (threshold effect), we did not report the individual weighted average for sensitivity, specificity, positive likelihood ratio, and negative likelihood ratio, as this would lead to under-estimation of diagnostic test performance (25,26). Instead, a summary receiver operator characteristic (sROC) curve of all the studies was created (27), as this is a better summary of the study results than a single joint summary estimate of sensitivity and specificity (see Appendix for details about construction and interpretation of sROC curve). We used the area under the sROC curve to judge the degree of accuracy of the tests according to published guidelines (28) (≥ 0.97 = excellent, 0.93 to 0.96 = very good, 0.75 to 0.92 = good, 0.50 to 0.75 = poor). We computed a weighted average of the specificity from all studies using the random effects models; then the sensitivity was calculated from the sROC curve equations. Positive and negative likelihood ratios were derived from the summary sensitivity and specificity.

Heterogeneity was described using the I^2 statistic (29) for pooling sensitivity, specificity, and positive and negative likelihood ratios. Sensitivity analyses were performed to evaluate the robustness of results according to blinding of test results among anesthesiologists (blinding versus unclear/no blinding) for the primary outcomes.

Publication bias was assessed using the Egger's weighted regression method (30) with precision (1/ standard error) and log odds ratio plotted. The intercept value in Egger's regression method provides an estimate of asymmetry of the funnel plot, with positive values indicating a trend towards higher levels of test accuracy in studies with smaller sample sizes. The threshold of significance was set at P < 0.10 for this method as this test has low power (31). All statistical analyses were performed using Stata version 8.0 (Stata

Corp, College Station, TX) and MetaDiSc version 1.1.1 (Zamora J, Muriel A, Abraira V, Madrid, Spain).

Results

Our literature search identified 42 studies that enrolled 34,513 patients (2,4,6–9,11,12,32–65). One study was excluded because of inconsistencies in the presented data (58). Two studies were excluded in which the modified Mallampati test was assessed as part of a more comprehensive risk score (53,57) but data for the modified Mallampati test component were unavailable. The characteristics of the included studies are summarized in Tables 2 and 3. There were no studies in children.

The quality of the studies was assessed according to the method of patient recruitment and blinding of the Mallampati tests and reference tests results among anesthesiologists. Patients were recruited consecutively in 19 studies (2,4,6,8,34,35,40–44,46,47,49,50,52,59,60,62). Constantikes (39) took a convenient sample of 30 patients. Blinding was reported in 10 studies (7,11,35,38,46,48–50,52,53). Both consecutive patient recruitment and blinding occurred in 5 studies (35,46,49,50,52).

There was poor documentation about how the Mallampati tests were done with regard to body and head positions and the use of phonation. Nine studies had adequate description of all three aspects (4,7,43,51,53,56,57,64,65). Phonation during the Mallampati test was described in six studies (33,39,42,44,53,65). The modified Mallampati test was performed in Asian patients in eight studies (8,11,12,46,55,56,59,64). The modified Mallampati test was used in obstetric patients in five studies (7,11,43,46,59). In two studies (11,59) separate data of test characteristics were given for obstetric and gynecological patients. As there was a discrepancy between the abstract and text in the proportion of obstetric and non-obstetric patients in one study (52), it was excluded from meta-regression analyses.

Difficult Laryngoscopy

All studies used Cormack and Lehane's original classification for defining difficult laryngoscopy except one study (55) that used the modified five-grade score (21).

The original Mallampati test was used in 9 studies with 14,438 patients (Table 2). The prevalence of difficult laryngoscopy ranged from 6% to 27%. There was a high prevalence of difficult laryngoscopy in patients with cervical disease (36) and in patients with diabetes (40). In one study, the authors attributed the high prevalence of difficult laryngoscopy to the use of the McCoy laryngoscope (37). The sensitivity and specificity of the individual studies ranged from 0.05 to 1.00

			4 22	Prevalence of difficult airway according								
Study	Patients (n)	Women, (%)	mean (range) (yr)	reference standard (%)	True positive (n)	False positive (n)	False negative (n)	True Negative (n)	Sensi- tivity	Speci- ficity	Positive Likihood Ratio	Negative Likelihood Ratio
Difficult laryn	goscopy											
Bilgin et al., 1998 (32)	500	51	46 (16-80)	9	39	12	7	442	0.85	0.97	32.08	0.16
Calder et al., 1995 (36)	251	51	54 (8-84)	20	29	8	22	192	0.57	0.96	14.22	0.45
Cohen et al., 1992 (33)	663	NS	NS (18-88)	6	21	55	29	558	0.42	0.91	4.68	0.64
Constantikes, 1993 (39)	30	NS	NS (18-NS)	7	2	8	0	20	1.00	0.71	2.84	0.24
El-Ganzouri et al., 1996 (6)	10507	NS	NS (18-NS)	6	287	1080	355	8785	0.45	0.89	4.08	0.62
Nadal et al., 1998 (40)	83	64	53 (18-NS)	27	1	1	21	60	0.05	0.98	2.77	0.97
Randell et al., 1998 (37)	100	38	48 (NS)	20	3	7	17	73	0.15	0.91	1.71	0.93
Tse et al., 1995 (35)	471	53	NS (19-89)	13	41	145	21	264	0.66	0.65	1.87	0.53
Voyagis et al., 1998 (34)	1833	NS	NS	8	134	227	18	1454	0.88	0.87	6.53	0.14
Difficult intub	ation											
Bilgin et al., 1998 (32)	500	51	46 (16-80)	8	17	34	23	426	0.43	0.93	5.75	0.62
Cohen et al., 1992 (33)	663	NS	NS (18-88)	11	25	51	49	538	0.34	0.91	3.90	0.73
El-Ganzouri et al., 1996 (6)	10507	NS	NS (18-NS)	6	287	1080	355	8785	0.45	0.89	4.08	0.62
Mallampati et al., 1985 (2)	210	78	39 (18-82)	13	14	1	14	181	0.50	1.00	91.0	0.50
Tse et al., 1995 (35)	471	53	NS (19-89)	13	41	145	21	264	0.66	0.65	1.87	0.53
Difficult ventil	lation											
El-Ganzouri et al., 1996 (6)	10507	NS	NS (18-NS)	0.1	3	1364	5	9135	0.38	0.87	2.89	0.72

Table 2. Prospective Studies of Original Mallampati Test with Three Reference Tests for Difficult Airway

NS = not specified.

and 0.65 to 0.98, respectively (Table 2). The positive and negative likelihood ratios of the 9 studies ranged from 1.71 to 32.08 and 0.14 to 0.97, respectively (Table 2). As there was an apparent relationship between sensitivity and specificity (Spearman r = 0.45), a sROC curve was constructed (Fig. 1). The area (\pm sE) under the symmetrical sROC curve was 0.89 (0.05). Considering the threshold effect, the diagnostic odds ratio (DOR) was 19.57 (95% CI, 5.02 to 76.27). The summary estimate for sensitivity was 0.71 derived from equation (2) for the sROC curve at the summary specificity of 0.89. Hence, the summary positive and negative likelihood ratios were 6.45 and 0.33, respectively. Blinding (relative DOR) (rDOR) 0.49; 95% confidence interval [CI], 0.02 to 12.72) and phonation (rDOR 0.11, 95% CI, 0.00 to 5.47) did not change the diagnostic performance of the original Mallampati test for predicting difficult laryngoscopy. There was no evidence of publication bias being present in this meta-analysis (t = 1.19, P = 0.27).

The modified Mallampati tests (3,4) were used in 19 studies with 10,579 patients (Table 3). There was wide variability in the prevalence of difficult laryngoscopy, which ranged from 2% to 26% (Table 2). The highest prevalence was in acromegaly patients (51). The sensitivity and specificity of the individual studies ranged from 0.12 to 1.00 and 0.44 to 0.98, respectively (Table 2). There was an association between sensitivity and specificity (Spearman's r = 0.32). The area (± sE) under the symmetrical sROC curve (Fig. 2) was 0.78 (0.05). Considering the threshold effect, the DOR was 6.45 (95% CI, 2.73 to 15.22). The summary estimate for sensitivity was 0.55 and the summary specificity was 0.84. The summary positive and negative likelihood ratios were 3.44 and 0.54, respectively. There was no significant difference in the areas under the sROC curve between the original and modified versions of the Mallampati test (z = 1.56; P = 0.12). Publication bias was not evident in the meta-analysis of 19 studies for difficult laryngoscopy (t = -0.57; P = 0.57).

Study, Year (Reference)	Patients (n)	Women, (%)	Age mean (range) (yr)	Prevalence of difficult airway according to reference standard (%)	True posi- tive (n)	False posi- tive (n)	False nega- tive (n)	True nega- tive (n)	Sensi- tivity	Speci- ficity	Positive Liki- hood Ratio	Negative Likeli- hood Ratio
Difficult laryngoscopy	r											
Adnet et al., 2001 (41) Bouaggard et al., 2004	1171 320	57 88	NS (15-90) NS (18-NS)	9 10	25 9	53 6	80 23	1013 282	0.24 0.28	0.95 0.98	4.79 13.50	0.80 0.73
(42) Brodsky et al., 2002	100	78	44 (NS)	9	5	28	4	63	0.56	0.69	1.81	0.64
(++) Butler et al 1992 (12)	220	61	NS (16-80)	8	10	38	8	164	0.56	0.81	2 95	0.55
Cattano et al., 2004 (9)	1956	NS	NS (18-NS)	2	15	182	28	1731	0.35	0.91	3.67	0.72
Constantikes, 1993 (39)	30	NS	NS (18-NS)	7	2	15	0	13	1.00	0.46	1.56	0.36
Ezri et al., 2001 (4)	764	48	44 (18-NS)	11	68	196	13	487	0.84	0.71	2.93	0.23
Ezri et al., 2003 (60)	1472	49	NS (18-NS)	10	116	354	36	966	0.76	0.73	2.85	0.32
Ezri et al., 2003 (61)	50	58	NS (NS)	18	3	6	6	35	0.33	0.85	2.28	0.78
Ezri et al., 2003 (62)	644	43	NS (40-NS)	7	11	208	32	393	0.26	0.65	0.74	1.14
Gercek et al., 2003 (38)	500	63	NS (NS)	5	3	14	22	461	0.12	0.97	4.07	0.91
Gupta et al., 2003 (46)	372	100	25 (16-37)	7	15	8	10	339	0.60	0.98	26.03	0.41
Ita et al., 1994 (45)	57	68	31 (NS)	5	2	2	1	52	0.67	0.96	18.00	0.35
Koh et al., 2002 (55)	605	56	NS (16-NS)	5	14	45	17	529	0.45	0.92	5.76	0.60
Merah et al., 2004 (43)	80	100	31 (NS)	10	7	3	1	69	0.88	0.96	21.00	0.13
Rocke et al., 1992 (7)	1500	100	26 (NS)	2	16	381	11	1092	0.59	0.74	2.29	0.55
Schmitt et al., 2000 (51)	128	53	46 (19-78)	26	25	53	8	42	0.76	0.44	1.36	0.55
Vani et al., 2000 (64)	50	56	57 (18-NS)	16	1	4	7	38	0.13	0.91	1.31	0.97
Yeo et al., 1992 (59) Difficult intubation	560	100	NS (15-69)	2	3	25	8	524	0.27	0.95	5.99	0.76
Adnet et al., 2001 (41)	1171	57	NS (15-90)	8	24	54	66	1027	0.27	0.95	5.34	0.77
Arne et al., 1998 (47)	1200	NS	NS (15-NS)	4	39	168	11	982	0.78	0.85	5.34	0.26
Avuso et al., 2003 (50)	181	26	54 (NS)	30	29	26	25	101	0.54	0.80	2.62	0.58
Bouaggard et al., 2004 (42)	320	88	NS (18-NS)	5	7	8	10	295	0.41	0.97	15.60	0.60
Brodsky et al., 2002 (44)	100	78	44 (NS)	12	7	26	5	62	0.58	0.71	1.97	0.59
Cattano et al., 2004 (9)	1956	NS	NS (18-NS)	3	28	169	28	1731	0.50	0.91	5.62	0.55
Frerk et al., 1991 (54)	244	41	44 (18-82)	5	9	43	2	190	0.82	0.82	4.43	0.22
Gercek et al., 2003 (38)	500	63	NS (NS)	7	2	15	31	452	0.06	0.97	1.89	0.97
Gupta et al., 2003 (46)	372	100	25 (16-37)	7	15	8	9	340	0.63	0.98	27.19	0.38
Iohom et al., 2003 (48)	212	49	NS (18-NS)	9	8	22	12	170	0.40	0.89	3.49	0.68
Ita et al., 1994 (45)	57	68	31 (NS)	2	0	4	1	52	0	0.93	3.17	0.81
Juvin et al., 2003 (63)	263	72	NS (18-79)	9	20	75	3	165	0.87	0.69	2.78	0.19
Kaul et al., (56)	500	42	36 (18-87)	8	31	24	11	434	0.74	0.95	14.09	0.28
Khan et al., 2003 (49)	300	NS	NS (16-NS)	6	14	94	3	189	0.82	0.67	2.48	0.26
Merah et al., 2004 (43)	80	100	31 (NS)	10	7	3	1	69	0.88	0.96	21.00	0.13
Rocke et al., 1992 (7)	1500	100	26 (NS)	2	19	378	13	1090	0.59	0.74	2.31	0.55
Savva, 1994 (52)	350	57	NS (18-81)	5	11	113	6	220	0.65	0.66	1.91	0.53
Wong et al., 1999 (11)	411	100	NS (NS)	2	6	150	1	254	0.86	0.63	2.29	0.23
Yamamoto et al., 1997 (8)	3680	49	52 (NS)	2	38	1723	18	1901	0.68	0.53	1.43	0.61
Yeo et al., 1992 (59) Difficult ventilation	560	100	NS (15-69)	2	3	25	8	524	0.27	0.95	5.99	0.76
Cattano et al., 2004 (9)	1956	NS	NS (18-NS)	2	15	179	32	1703	0.32	0.91	3.36	0.75
Langeron et al., 2000 (65)	1502	55	NS (18-NS)	5	17	186	58	1241	0.23	0.87	1.74	0.89

Table 3. Prospective Studies of Modified Mallampati Test with Three Reference Tests for Difficult Airway

NS = not specified.

Blinding (rDOR 1.49; 95% CI, 0.22 to 10.31) and phonation (rDOR 0.86; 95% CI, 0.13 to 5.76) did not change the diagnostic performance of the modified Mallampati test for predicting difficult laryngoscopy.

Also, there were no differences in diagnostic test performance between studies of Asian and Caucasian patients (rDOR 1.09; 95% CI, 0.21 to 5.64). However, on meta-regression, the modified Mallampati test was



Figure 1. Summary receiver operating characteristic curve of the original Mallampati test for predicting difficult laryngoscopy. Each circle represents the results of a single study. The summary estimates of sensitivity and specificity are 0.71 and 0.89, respectively.

5.08 (95% CI, 1.26 to 20.58; P = 0.03) times more accurate in studies of obstetric patients than in studies in surgical patients.

Difficult Tracheal Intubation

There was wide variation in the definition of difficult tracheal intubation. Many studies (6,8,11,35,43,49,56,59) used the original Cormack and Lehane definition (20). Three studies (41,42,63) defined difficult tracheal intubation as a score >5 on the Intubation Difficulty Score described by Adnet et al. (41). This scoring system incorporates the number of attempts, number of additional operators, number of alternative intubation techniques, Cormack and Lehane grade, lifting force, laryngeal pressure, and vocal cord mobility. Two studies (7,46) in obstetric patients used Rocke et al.'s classification (7) for defining difficult tracheal intubation.

The original Mallampati test (2) was used to predict difficult tracheal intubation in 5 studies enrolling 12,351 patients (Table 2). The prevalence of difficult tracheal intubation ranged from 6% to 13%. There were low sensitivities (0.34 to 0.66) and varying specificities (0.65 to 1.00). The positive likelihood ratios varied from 1.87 to 91.0; negative likelihood ratios varied from 0.50 to 0.73. There was an association between sensitivity and specificity (Spearman's r = 0.30). As the DOR was not constant across the threshold (b = -0.71; 95% CI, -1.21 to -0.21), the sROC curve was asymmetrical (Fig. 3). The area (\pm sE) under



Figure 2. Summary receiver operating characteristic curve of the modified Mallampati test for predicting difficult laryngoscopy. The summary estimates of sensitivity and specificity are 0.55 and 0.84, respectively.

the asymmetrical sROC was 0.58 (0.12). The summary estimate for the sensitivity was 0.50 derived from the equation (3) for the sROC curve at the summary specificity of 0.89. Phonation during the test (rDOR 0.27; 95% CI, 0.00 to 15.55) and blinding (rDOR 15.65; 95% CI, 0.14 to 1784.76) did not appear to affect the overall accuracy of the test. There was no evidence of publication bias being present in the studies pooled (t = -0.39, P = 0.72).

Twenty studies enrolling 13,957 patients (Table 3) examined the use of the modified Mallampati (3,4) test for predicting difficult tracheal intubation. The prevalence of difficult tracheal intubation ranged from 2% to 30% (Table 3). The high prevalence of difficult tracheal intubation (30%) occurred in patients with pharyngolaryngeal disease (50). The sensitivities ranged from 0 to 0.88 and the specificities ranged from 0.53 to 0.98 (Table 3). The positive likelihood ratios varied 1.43 to 27.19; negative likelihood ratios varied from 0.13 to 0.97 (Table 3). There was a correlation between sensitivity and specificity (Spearman's r = 0.47). The area (\pm sE) under the symmetrical sROC (Fig. 4) was 0.83 (0.03). Considering the threshold effect, the DOR was 10.43 (95% CI, 5.32 to 20.48). The summary estimate for sensitivity was 0.76 when the summary specificity was 0.77. Hence, the summary positive and negative likelihood ratios were 3.30 and 0.31, respectively. The modified Mallampati test was better at identifying difficult tracheal intubation than the original Mallampati test (z = 2.02; P = 0.04). There was



Figure 3. Summary receiver operating characteristic curve of the original Mallampati test for predicting difficult tracheal intubation. Each circle represents the results of a single study. The summary estimates of sensitivity and specificity are 0.50 and 0.89, respectively.

no evidence of publication bias in the 20 studies pooled for difficult tracheal intubation (t = 0.89; P = 0.39).

Blinding (rDOR 0.52; 95% CI, 0.17 to 1.65), phonation (rDOR 0.94; 95% CI, 0.13 to 6.69), and studies of Asian patients (rDOR 1.65; 95% CI; 0.44 to 6.24) did not change the diagnostic performance of the modified Mallampati test for predicting difficult tracheal intubation. The difference between obstetric patients and non-obstetric patients in the accuracy of the modified test for predicting difficult tracheal intubation was not significant (rDOR 2.69; 95% CI, 0.81 to 8.88, P = 0.10).

No failed intubation occurred in 12 studies (4,12,31,34,39-41,46-48,50,62). When failed intubation occurred, the prevalence varied from 0.1% (7) to 3.8% (57).

Difficult Ventilation

Three studies recorded difficult ventilation (6,9,65). Definitions varied and included inability to obtain chest excursion sufficient to maintain a clinically acceptable capnogram waveform despite optimal head and neck positioning, use of muscle paralysis, use of an oral airway, and optimal application of a facemask (6). In another study (65), ventilation via a mask was considered difficult only when the anesthesiologist considered that the difficulty was clinically relevant and could have led to problems if mask ventilation



Figure 4. Summary receiver operating characteristic curve of the modified Mallampati test for predicting difficult tracheal intubation. The summary estimates of sensitivity and specificity are 0.76 and 0.77, respectively.

had to be maintained for a longer time. Bag-mask ventilation was considered difficult if one or more of the following factors were present: inability to maintain an adequate seal; inability to obtain chest excursion, obtain a good capnograph tracing, or maintain oxygen saturation more than 90% despite good muscle relaxation; the necessity of using a Guedel oral airway; or two-person bag-mask ventilation (9). The sensitivity, specificity, and positive and negative likelihood ratios are shown in Tables 2 and 3 for this outcome. Pooled sensitivity and specificity of the modified Mallampati test (3) were 0.26 (95% CI, 0.19 to 0.35) and 0.89 (95% CI, 0.88 to 0.90), respectively (9,65). There was little heterogeneity between the two studies for sensitivity ($I^2 = 21\%$). However, there was substantial heterogeneity for specificity ($I^2 = 90\%$). The positive and negative likelihood ratios were 2.42 (95% CI, 1.25 to 4.66) and 0.83 (95% CI, 0.71 to 0.98), respectively, suggesting poor accuracy. There were moderate amounts of heterogeneity among studies for positive and negative likelihood ratios ($I^2 = 78\%$ and 52% respectively).

To put the results of this systematic review in a clinical context, readers can estimate the post-test probability of a difficult airway after an examination of the airway using the modified Mallampati test according to the prevalence of difficult airway in their population (Fig. 5). The range of pre-test probabilities reflects the range of prevalence reported in this systematic review. If the pre-test probability of difficult airway is 10%, a positive test generates a post-test



Figure 5. Post-test probability of difficult laryngoscopy and difficult tracheal intubation after using the modified Mallampati test. Post-test probabilities are shown as a function of pretest probability for positive (Grade III or IV) and negative results (Grade 0, I or II) on the modified Mallampati test.

probability of difficult laryngoscopy of 28% and difficult intubation of 27%; a negative test generates a post-test probability of difficult laryngoscopy of 6% and difficult intubation of 3%.

Discussion

This systematic review of the literature identified many studies describing the performance of the original (2) and modified (3,4) Mallampati tests to predict the difficult airway. There was substantial variability in the reported sensitivity and specificity among the studies and in definitions of the reference tests. Unlike meta-analysis of interventions, which produces one answer, the performance of diagnostic tests is affected by changes in sensitivity and specificity as reflected in the sROC curves (Figs. 1 to 4). Thus, there is no unique joint summary estimate of sensitivity and specificity; it is only possible to obtain a summary estimate of one value conditional on the value of the other (66). Overall, the accuracy of the Mallampati tests was poor to good, depending on the version of the test and reference test used. Our results are not directly comparable to a recent meta-analysis of bedside screening test for predicting difficult intubation (67). In Shiga et al.'s meta-analysis (67), there was no distinction made between the various versions of the Mallampati test or between difficult laryngoscopy and difficult intubation, a major limitation of their study. Nevertheless, they concluded that the Mallampati test's clinical value of a bedside screening test was limited as it had poor to moderate discriminative power when used alone. Our results concur with this view.

Both versions of the Mallampati test had good accuracy for identifying difficult laryngoscopy as assessed according to the original and modified Cormack and Lehane grading system. This system is widely used in clinical practice to describe the best view obtained by direct laryngoscopy with or without manipulation of the larynx. However, there is considerable uncertainty and inaccuracy in this grading system, especially between grade 2 and grade 3 (68). The incidence of difficult laryngoscopy may be underestimated, as most of the studies used the original Cormack and Lehane grading system. Approximately 3% (55) to 7% (21) of patients graded 2b, who would otherwise have been rated grade 2 in the original system, will have a high risk of difficult laryngoscopy. Such misclassification may affect the overall test performance of the Mallampati tests. Many studies used the same Cormack and Lehane grading system to define both difficult laryngoscopy and difficult intubation. Although difficult intubation is the end result of difficult laryngoscopy, the former also depends on the operator's experience, patient characteristics, and clinical setting.

The recommended best way to perform the Mallampati test for predicting difficult laryngoscopy is putting the patient in a sitting position, with the head in full extension, the tongue out, and with phonation (53). However, many studies did not specifically document the way the Mallampati test was performed. Therefore, variations in the conduct of Mallampati tests may contribute to some of the heterogeneity of results seen in this systematic review. Unexpectedly, phonation did not influence the overall accuracy of the Mallampati tests.

There was a large variation among studies in the definition of difficult tracheal intubation. There is no current consensus on the definition of difficult tracheal intubation. Therefore, we used the definition from each study to establish an operational reference standard reflecting current clinical practice. The different definitions of difficult tracheal intubation may explain, in part, the heterogeneity of results in the sROC curves. For predicting difficult tracheal intubation, the original Mallampati test had very poor accuracy. Four of the five studies had sensitivities <50%. Small increases in sensitivity led to large sacrifices in specificity. The asymmetrical sROC curve suggests that accuracy was dependent on threshold. The lowest accuracy occurred when the threshold was high. This may be related to the quality of study. The lowest accuracy occurred in a study with the least amount of reviewer and patient selection bias (35). In contrast, the modified Mallampati test had good accuracy for predicting difficult tracheal intubation and was significantly better than the original test. The discrepancy in results between the two versions of the Mallampati test may be related to the definition of difficult tracheal intubation used and difference in the study populations.

The accuracy of the modified Mallampati test for predicting difficult laryngoscopy was five times higher in obstetric patients than in non-obstetric patients, although for predicting difficult tracheal intubation, the difference was not significant. This is consistent with studies that showed that pregnancy caused a 34% increase in Mallampati grade 4 (10) and that the risk of difficult intubation in obstetric patients was approximately 8 times more than in surgical patients (3). More difficult laryngoscopy in obstetric patients most likely occurs because of facial and pharyngeal edema secondary to hormonally induced fluid retention (69). These results suggest that the Mallampati tests are probably better at predicting difficult laryngoscopy associated with soft tissue changes compared with other anatomical factors.

We found little evidence of ethnic differences in the accuracy of modified Mallampati tests for difficult laryngoscopy and difficult intubation, despite known cephalometric differences among ethnic populations (70).

In a recent editorial, Murphy et al. (71) suggested that we should focus on "ventilatability" rather than "intubatability." The accuracy of the Mallampati tests for predicting difficult mask ventilation was poor, but this was based on only three studies. Therefore, these results should be interpreted with caution. For predicting difficult mask ventilation, the presence of 2 of 5 factors (age older than 55 years, body mass index $>26 \text{ kg/m}^2$, lack of teeth, presence of beard, and history of snoring) was associated with good accuracy (area under the curve 0.76 ± 0.11) (65). As expected, there was a strong association between difficult tracheal intubation and difficult mask ventilation (65).

Systematic review and meta-analysis are considered to provide the least biased estimates of effect but if the "raw material" is flawed, then the conclusions of systematic reviews will be compromised and invalid (66). The quality of reporting varied among studies; only a few studies described study methodology and Mallampati test assessments in adequate detail. We assumed that the quality of the study was inadequate if it was clearly stated that there were deficiencies in design and conduct. Omission of reporting specific details of a study was associated with systematic differences in results (72). Of the 42 studies included in this systematic review, only 5 studies recruited patients consecutively with test results blinded among anesthesiologists. This suggests that the majority of studies included in this systematic review may have

less than adequate study methodology. Future studies of tests for identifying difficult airway should adopt the Standards for Reporting of Diagnostic Accuracy guidelines (73). This would allow readers to assess the potential for bias in the study and to evaluate the generalizability of study results.

Interpreting the reference test with knowledge of the results of the test under study can lead to an over-estimation of a test's accuracy (72). This is known as review bias. Unblinded studies tend to overestimate the diagnostic test accuracy by 1.3 times (95% CI, 1.0 to 1.9) (72). However, we did not find a significant effect of blinding on the Mallampati tests' accuracy. We also minimized spectrum bias (study sample does not include the complete spectrum of patient characteristics) by excluding case-control studies from the systematic review. Diagnostic accuracy can be overestimated by 3 times (95% CI, 2.0 to 4.5) if the test is evaluated in a group of patients already known to have the disease and a separate group of normal patients, as in case-control studies (72).

Publication bias in meta-analyses of test accuracy is highly prevalent (22). This type of bias is a threat to the validity of meta-analysis as it can lead to inappropriate decision making and health care policies. We undertook a comprehensive literature search using several electronic databases. Although we restricted our systematic review to include English language studies, the inclusion of non-English language studies would only increase the precision without affecting the overall accuracy estimates. A previous study showed no relationship between publication bias and language restriction in reviews (22). We believe that our results are robust, as publication bias was not present.

As there were no pediatric studies, the results of our systematic review are applicable only to adults. There was a wide range of difficult airway prevalence, reflecting various patient characteristics, including pregnancy (7,11,43,46,59), pharyngolaryngeal disease (50), acromegaly (51), and obesity (44,61). As post-test probability depends on the disease prevalence, knowledge of the prevalence of difficult laryngoscopy and difficult intubation at any individual hospital will aid in the application of our results (Fig. 5). The decision to perform additional radiographic evaluation, consultation with other specialists or use special techniques/ equipment to manage difficult airways will depend on how high the post-test probability is and, consequently, at what level the treatment threshold is set by the individual anesthesiologist.

The results of our systematic review question the routine use of the Mallampati tests. Given the poor to moderate inter-observer reliability of the modified Mallampati test (74,75) and the poor to good accuracy of the Mallampati tests, should we abandon their use? To decide this, anesthesiologists should balance the cost of failing to predict a difficult airway when there is a false negative result versus the possibility of unnecessary treatment when there is a false positive result. Used alone, the Mallampati tests are insufficient to confidently predict the presence or absence of a difficult airway; we believe they should form only a limited part of the overall assessment of the airway. As recommended by the American Society of Anesthesiologists Task Force on the management of the difficult airway (23), dentition, thyromental distance, and neck extension are other parts of the airway examination that also need to be examined.

Appendix

The sROC curve method considers heterogeneity across studies attributable to differences in the threshold values used. Even if the same threshold has been used, inter-observer differences in the Mallampati grades (74,75) may lead to inherent variations in the positive results cutoff. To confirm that there was a threshold effect, the true-positive rate (TPR) and falsepositive rate (FPR) of each study were plotted against each other, and the Spearman correlation coefficient was calculated. In creating the sROC, the TPR and FPR were converted to their logits, and the sum and differences of the logits were estimated. Equally unweighted least squares linear regression of the following model was performed:

$$D = a + bS \tag{1}$$

where D = logit TPR - logit FPR, S = logit TPR + logitFPR, a = intercept term, and b = regression coefficient for S. D is equivalent to the diagnostic odds ratio (DOR), which conveys the test's accuracy in discriminating diseased subjects from nondiseased subjects (76). S can be interpreted as a measure of the diagnostic test threshold, with high values corresponding to liberal inclusion criteria for diseased subjects (76). The regression coefficient *b* represents the dependence of the test accuracy on threshold. If $b \approx 0$, then the studies are homogeneous and can be summarized by an overall DOR noting that $a = \ln(\text{DOR})$ (76), giving a symmetrical sROC. The studies are heterogeneous with respect to the diagnostic odds ratio if $b \neq 0$ (76). In this case, the sROC is asymmetrical. The DOR is related to the area under the sROC curve. A DOR of 1 is equivalent to an area under the sROC curve of 50%; the larger the DOR, the larger the area under the sROC curve. Ninety-five percent confidence intervals (95% CI) were estimated around the DOR. The areas under the sROC curve of the original and modified Mallampati test were compared using the method outlined by Hasselblad and Hedges (77). The resulting equations below (equations 2 and 3) represent the logit form of the sROC curve, from which a pooled estimate of TPR and FPR can be obtained. The equation (66) of the corresponding symmetrical sROC curve is given by:

$$sensitivity = \frac{1}{1 + \frac{1}{DORx\left(\frac{1 - specificity}{specificity}\right)}}$$
(2)

The equation (26) of the corresponding asymmetrical sROC curve is given by:

$$sensitivity = \frac{1}{1 + \frac{1}{e^{a/(1-b)}x\left(\frac{1 - specificity}{specificity}\right)^{(1+b)/(1-b)}}}$$
(3)

Meta-regression was used to explore possible reasons for heterogeneity with *a priori* subgroups, including type of patient population (coded 1 = Asians, 0 =Caucasians; 1 = obstetrics, 0 = non-obstetrics) and phonation (coded 1 = yes, 0 = no) during the Mallampati tests. This was done by extending the sROC model introduced above (equation 1) to include a covariate (27). The resulting parameter estimates of the covariate can be interpreted, after antilogarithm transformation, as the relative DOR (rDOR) (72) and reflects the differences in threshold choice at different levels of the covariate. Fitting a covariate to the model does not result in a separate sROC curve for each level of the covariate, as the relationship between TPR and FPR is reflected only in *a* (78).

The authors thank the authors of the original studies who responded to our requests for unpublished and additional data.

References

- Crosby ET, Cooper RM, Douglas MJ, et al. The unanticipated difficult airway with recommendations for management. Can J Anaesth 1998;45:757–76.
- 2. Mallampati SR, Gatt SP, Gugino LD, et al. A clinical sign to predict difficult tracheal intubation: a prospective study. Can Anaesth Soc J 1985;32:429–34.
- Samsoon GL, Young JR. Difficult tracheal intubation: a retrospective study. Anaesthesia 1987;42:487–90.
- 4. Ezri T, Warters RD, Szmuk P, et al. The incidence of class "zero" airway and the impact of Mallampati score, age, sex, and body mass index on prediction of laryngoscopy grade. Anesth Analg 2001;93:1073–5.
- 5. Mellado PF, Thunedborg LP, Swiatek F, Kristensen MS. Anaesthesiological airway management in Denmark: assessment, equipment and documentation. Acta Anaesthesiol Scand 2004; 48:350–4.

- 6. El-Ganzouri AR, McCarthy RJ, Tuman KJ, et al. Preoperative airway assessment: predictive value of a multivariate risk index. Anesth Analg 1996;82:1197–204.
- Rocke DA, Murray WB, Rout CC, Gouws E. Relative risk analysis of factors associated with difficult intubation in obstetric anesthesia. Anesthesiology 1992;77:67–73.
- Yamamoto K, Tsubokawa T, Shibata K, et al. Predicting difficult intubation with indirect laryngoscopy. Anesthesiology 1997;86: 316–21.
- 9. Cattano D, Panicucci E, Paolicchi A, et al. Risk factors assessment of the difficult airway: an Italian survey of 1956 patients. Anesth Analg 2004;99:1774–9.
- 10. Pilkington S, Carli F, Dakin MJ, et al. Increase in Mallampati score during pregnancy. Br J Anaesth 1995;74:638–42.
- 11. Wong SH, Hung CT. Prevalence and prediction of difficult intubation in Chinese women. Anaesth Intensive Care 1999;27: 49–52.
- 12. Butler PJ, Dhara SS. Prediction of difficult laryngoscopy: an assessment of the thyromental distance and Mallampati predictive tests. Anaesth Intensive Care 1992;20:139–42.
- 13. Deville WL, Buntinx F, Bouter LM, et al. Conducting systematic reviews of diagnostic studies: didactic guidelines. BMC Med Res Methodol 2002;2:9.
- Irwig L, Tosteson AN, Gatsonis C, et al. Guidelines for metaanalyses evaluating diagnostic tests. Ann Intern Med 1994;120: 667–76.
- Sonbul ZM, Demian AD, Deiab DG, El Asfour AA. Prediction of difficult intubation in obstructive sleep apnoea patients. Egyptian J Anaesth 2004;20:189–94.
- Naguib M, Malabarey T, AlSatli RA, et al. Predictive models for difficult laryngoscopy and intubation: a clinical, radiologic and three-dimensional computer imaging study. Can J Anaesth 1999;46:748–59.
- Frerk CM, Till CB, Bradley AJ. Difficult intubation: thyromental distance and the atlanto-occipital gap. Anaesthesia 1996;51: 738–40.
- Samra SK, Schork MA, Guinto FC Jr. A study of radiologic imaging techniques and airway grading to predict a difficult endotracheal intubation. J Clin Anesth 1995;7:373–9.
- 19. Karkouti K, Rose DK, Wigglesworth D, Cohen MM. Predicting difficult intubation: a multivariable analysis. Can J Anaesth 2000;47:730–9.
- 20. Cormack RS, Lehane J. Difficult tracheal intubation in obstetrics. Anaesthesia 1984;39:1105–11.
- 21. Yentis SM, Lee DJ. Evaluation of an improved scoring system for the grading of direct laryngoscopy. Anaesthesia 1998;53: 1041–4.
- Song F, Khan KS, Dinnes J, Sutton AJ. Asymmetric funnel plots and publication bias in meta-analyses of diagnostic accuracy. Int J Epidemiol 2002;31:88–95.
- 23. 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.
- 24. Jaeschke R, Guyatt GH, Sackett DL. Users' guides to the medical literature. III. How to use an article about a diagnostic test. B. What are the results and will they help me in caring for my patients? The Evidence-Based Medicine Working Group. JAMA 1994;271:703–7.
- Shapiro DE. Issues in combining independent estimates of the sensitivity and specificity of a diagnostic test. Acad Radiol. 1995;2(Suppl 1):S37–47.
- 26. Littenberg B, Moses LE. Estimating diagnostic accuracy from multiple conflicting reports: a new meta-analytic method. Med Decis Making 1993;13:313–21.
- Moses LE, Shapiro D, Littenberg B. Combining independent studies of a diagnostic test into a summary ROC curve: dataanalytic approaches and some additional considerations. Stat Med 1993;12:1293–316.
- Jones CM, Athanasiou T. Summary receiver operating characteristic curve analysis techniques in the evaluation of diagnostic tests. Ann Thorac Surg 2005;79:16–20.

- 29. Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. BMJ 2003;327:557–60.
- Egger M, Davey SG, Schneider M, Minder C. Bias in metaanalysis detected by a simple, graphical test. BMJ 1997;315: 629–34.
- Macaskill P, Walter SD, Irwig L. A comparison of methods to detect publication bias in meta-analysis. Stat Med 2001;20: 641–54.
- Bilgin H, Ozyurt G. Screening tests for predicting difficult intubation: a clinical assessment in Turkish patients. Anaesth Intensive Care 1998;26:382–6.
- 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.
- 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.
- 35. 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.
- Calder I, Calder J, Crockard HA. Difficult direct laryngoscopy in patients with cervical spine disease. Anaesthesia 1995;50:756–63.
- Randell T, Maattanen M, Kytta J. The best view at laryngoscopy using the McCoy laryngoscope with and without cricoid pressure. Anaesthesia 1998;53:536–9.
- 38. Gercek A, Lim S, Isler FB, et al. The prediction of difficult intubation with bedside scoring systems. Marmara Med J 2003; 16:16–9.
- 39. Constantikes J. Predicting difficult tracheal intubation using a modified Mallampati sign: a pilot study report. CRNA 1993;4: 16–20.
- 40. Nadal JL, Fernandez BG, Escobar IC, et al. The palm print as a sensitive predictor of difficult laryngoscopy in diabetics. Acta Anaesthesiol Scand 1998;42:199–203.
- 41. Adnet F, Racine SX, Borron SW, et al. A survey of tracheal intubation difficulty in the operating room: a prospective observational study. Acta Anaesthesiol Scand 2001;45:327–32.
- Bouaggad A, Nejmi SE, Bouderka MA, Abbassi O. Prediction of difficult tracheal intubation in thyroid surgery. Anesth Analg 2004;99:603–6.
- Merah NA, Foulkes-Crabbe DJ, Kushimo OT, Ajayi PA. Prediction of difficult laryngoscopy in a population of Nigerian obstetric patients. West Afr J Med 2004;23:38–41.
- Brodsky JB, Lemmens HJM, Brock-Utne JG, et al. Morbid obesity and tracheal intubation. Anesth Analg 2002;94:732–6.
- 45. 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.
- 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.
- Arne J, Descoins P, Fusciardi J, et al. Preoperative assessment for difficult intubation in general and ENT surgery: predictive value of a clinical multivariate risk index. Br J Anaesth 1998;80: 140–6.
- 48. Iohom G, Ronayne M, Cunningham AJ. Prediction of difficult tracheal intubation. Eur J Anaesthesiol 2003;20:31–6.
- 49. 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.
- 50. Ayuso MA, Sala X, Luis M, Carbo JM. Predicting difficult orotracheal intubation in pharyngo-laryngeal disease: preliminary results of a composite index. Can J Anaesth 2003;50:81–5.
- Schmitt H, Buchfelder M, Radespiel-Troger M, Fahlbusch R. Difficult intubation in acromegalic patients: incidence and predictability. Anesthesiology 2000;93:110–4.
- 52. Savva D. Prediction of difficult tracheal intubation. Br J Anaesth 1994;73:149–53.

- 53. Lewis M, Keramati S, Benumof JL, Berry CC. What is the best way to determine oropharyngeal classification and mandibular space length to predict difficult laryngoscopy? Anesthesiology 1994;81:69–75.
- 54. Frerk CM. Predicting difficult intubation. Anaesthesia 1991;46: 1005–8.
- 55. 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.
- Kaul TK, Deepika PG Singh A, Bansal M. Prediction of difficult intubation: analysis in 500 adult patients. J Anaesthesiol Clin Pharmacol 1998;14:323–8.
- 57. Ayoub C, Baraka A, el Khatib M, et al. A new cut-off point of thyromental distance for prediction of difficult airway. Middle East J Anesthesiol 2000;15:619–33.
- 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.
- Yeo SW, Chong JL, Thomas E. Difficult intubation: a prospective study. Singapore Med J 1992;33:362–4.
- Ezri T, Medalion B, Weisenberg M, et al. Increased body mass index per se is not a predictor of difficult laryngoscopy. Can J Anaesth 2003;50:179–83.
- 61. Ezri T, Gewurtz G, Sessler DI, et al. Prediction of difficult laryngoscopy in obese patients by ultrasound quantification of anterior neck soft tissue. Anaesthesia 2003;58:1111–4.
- 62. Ezri T, Weisenberg M, Khazin V, et al. Difficult laryngoscopy: incidence and predictors in patients undergoing coronary artery bypass surgery versus general surgery patients. J Cardiothorac Vasc Anesth 2003;17:321–4.
- 63. Juvin P, Lavaut E, Dupont H, et al. Difficult tracheal intubation is more common in obese than in lean patients. Anesth Analg 2003;97:595–600.
- 64. 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.
- 65. Langeron O, Masso E, Huraux C, et al. Prediction of difficult mask ventilation. Anesthesiology 2000; 92:1229–36.

- Egger M, Davey SG, Altman DG. Systematic reviews in health care, 2nd ed. London: British Medical Journal Publishing Group, 2001.
- Shiga T, Wajima Z, Inoue T, Sakamoto A. Predicting difficult intubation in apparently normal patients: a meta-analysis of bedside screening test performance. Anesthesiology 2005;103: 429–37.
- 68. Cohen AM, Fleming BG, Wace JR. Grading of direct laryngoscopy: a survey of current practice. Anaesthesia 1994;49:522–5.
- Hawthorne L, Wilson R, Lyons G, Dresner M. Failed intubation revisited: 17-yr experience in a teaching maternity unit. Br J Anaesth 1996;76:680–4.
- Samman N, Mohammadi H, Xia J. Cephalometric norms for the upper airway in a healthy Hong Kong Chinese population. Hong Kong Med J 2003;9:25–30.
- Murphy M, Hung O, Launcelott G, et al. Predicting the difficult laryngoscopic intubation: are we on the right track? Can J Anaesth 2005;52:231–5.
- Lijmer JG, Bossuyt PM, Heisterkamp SH. Exploring sources of heterogeneity in systematic reviews of diagnostic tests. Stat Med 2002;21:1525–37.
- Bossuyt PM, Reitsma JB, Bruns DE, et al. Towards complete and accurate reporting of studies of diagnostic accuracy: the STARD initiative. BMJ 2003;326:41–4.
- 74. Karkouti K, Rose DK, Ferris LE, et al. Inter-observer reliability of ten tests used for predicting difficult tracheal intubation. Can J Anaesth 1996;43:554–9.
- Tham EJ, Gildersleve CD, Sanders LD, et al. Effects of posture, phonation and observer on Mallampati classification. Br J Anaesth 1992;68:32–8.
- Walter SD. Properties of the summary receiver operating characteristic (SROC) curve for diagnostic test data. Stat Med 2002; 21:1237–56.
- Hasselblad V, Hedges LV. Meta-analysis of screening and diagnostic tests. Psychol Bull 1995;117:167–78.
- Irwig L, Macaskill P, Glasziou P, Fahey M. Meta-analytic methods for diagnostic test accuracy. J Clin Epidemiol 1995;48: 119–30.