



ORIGINAL INVESTIGATION

Is Mallampati classification a good screening test? A prospective cohort evaluating the predictive values of Mallampati test at different thresholds

Clístenes C. de Carvalho^{a,b,*}, Danielle M. da Silva^c, Marina S. Leite^c, Flávia A. de Orange^{a,c}

^a Instituto de Medicina Integral Professor Fernando Figueira, Recife, PE, Brazil

^b Universidade Federal de Campina Grande, Campina Grande, PB, Brazil

^c Hospital das Clínicas de Pernambuco, Recife, PE, Brazil

Received 23 March 2020; accepted 5 September 2021

Available online 12 October 2021

KEYWORDS

Airway management;
Laryngoscopy;
Intubation,
intratracheal;
Sensitivity and
specificity

Abstract

Background: There is currently some discussion over the actual usefulness of performing preoperative upper airway assessment to predict difficult airways. In this field, modified Mallampati test (MMT) is a widespread tool used for prediction of difficult airways showing only a feeble predictive performance as a diagnostic test. We therefore aimed at evaluating if MMT test would perform better when used as a screening test rather than diagnostic.

Methods: An accuracy prospective study was conducted with 570 patients undergoing general anesthesia for surgical procedures. We collected preoperatively data on sex, age, weight, height, body mass index (BMI), ASA physical status, and MMT. The main outcome was difficult laryngoscopy defined as Cormack and Lahane classes 3 or 4. Bivariate analyses were performed to build three different predictive models with their ROC curves.

Results: Difficult laryngoscopy was reported in 36 patients (6.32%). Sex, ASA physical status, and MMT were associated with difficult laryngoscopy, while body mass index (BMI) was not. The MMT cut-off with the highest odds ratio was the class II, which also presented significantly higher sensitivity (94.44%). The balanced accuracy was 67.11% (95% CI: 62.78–71.44%) for the cut-off of class II and 71.68% (95% CI: 63.83–79.54) for the class III.

Conclusion: MMT seems to be more clinically useful when the class II is employed as the threshold for possible difficult laryngoscopies. At this cut-off, MMT shows the considerable highest sensitivity plus the highest odds ratio, prioritizing thus the anticipation of difficult laryngoscopies.

© 2021 Sociedade Brasileira de Anestesiologia. Published by Elsevier Editora Ltda. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

* Corresponding author.

E-mail: clistenescristian@hotmail.com (C.C. de Carvalho).

Introduction

Difficult airway management is one of the major causes of severe complications during general anesthesia.^{1–5} It makes the accurate prediction of difficult airways a desired faculty amongst the anesthesiologists as they would have more time to define the best managing strategy for such airways, bringing then more safety to the patients.⁶ Nevertheless, the ability to anticipate difficult airways remains fairly poor as no accurate predictor available so far presents great predictive performance.^{5–11} It consequently leads some professionals to question the actual benefit of performing preoperative upper airway assessment.⁸

In 1985, Mallampati et al. presented for the first time a bedside test that would posteriorly become one of the most widely used tools for prediction of difficult airways – the Mallampati classification.¹² From that time to the present moment, several studies have been conducted to better evaluate the predictive performance of both Mallampati and modified Mallampati¹³ (MMT) tests. Afterwards, an appraisal throughout these studies shows us a questionable predictive performance of the MMT and a disappointing ability of the test to segregate easy and difficult airways.^{6,14}

On the other hand, it is well accepted that a screening test should focus its predictive performance in hitting those patients with the positive outcome. To achieve this target, a screening test is supposed to have a great sensitivity, even if at the expense of a compromised specificity. This way, focusing on MMT as a screening test and hence choosing a lower threshold with highest sensitivity might improve its predictive performance and patients' safety.¹⁵

Therefore, we conducted this study to evaluate if a threshold other than the conventional Mallampati class III might be more clinically useful by performing better as a screening test and focusing on hitting difficult laryngoscopies.

Methods

This prospective study was performed at the surgical theater of the Federal University of Pernambuco's Teaching Hospital between October 2015 and January 2020. The study started only after the Ethical Committee approval and patients were only included if they agreed to participate and signed the informed consent form or the informed assent form in the case of patients under 18 years of age.

We consecutively included patients expected to undergo surgical procedures under general anesthesia, aging above 15 years. We excluded patients submitted to awake laryngoscopy and patients not submitted to direct laryngoscopy with a Macintosh blade. The variables evaluated were age; sex; height; weight; body mass index (BMI); American Society of Anesthesiologists (ASA) physical status; MMT; and difficult laryngoscopy.

Laryngoscopy was described as difficult when the airway manager (experienced 1 to 32 years), using direct laryngoscopy, classified the patient as 3a or higher according to Cormack and Lehane's classification system modified by Cook.¹⁶

A preanesthesia evaluation was performed just before surgeries assessing the MMT, determined with patient

seated, examiner's eyes at the level of patient's mouth, and patient with her/his mouth opened as widely as possible without phonation. The individual was then classified as Mallampati I, II, III, or IV.¹³ Posteriorly, three different cut-offs were used to indicate possible difficult laryngoscopies: test 1 (I = easy; II, III, and IV = difficult); test 2 (I and II = easy; III and IV = difficult); and test 3 (I, II, and III = easy; IV = difficult). The preanesthetic evaluation was performed by DdaS and ML, who were not involved in induction of anesthesia or laryngoscopies.

Following the intravenous access and the routine monitoring (i.e., ECG, pulse oximetry, and noninvasive blood pressure), the patients were put in sniffing position and pre-oxygenated. General anesthesia was induced according to the clinical judgment of the attending anesthesiologist. After neuromuscular blockade, a manual facemask ventilation was performed for 3–5 minutes, and laryngoscopies and tracheal intubations were then carried using the appropriate Macintosh blade by one of the residents or attending anesthesiologists. No measure of the depth of neuromuscular blockade was performed. The degree of glottic visualization was then recorded from the best view achieved without external laryngeal manipulation according to the Cormack and Lehane classification system modified by Cook.¹⁶

Statistical analysis

The data analysis was performed using the R project software program.^{17–21} First, a descriptive analysis was performed, with percentages and measures of central tendency and dispersion (means and standard deviation, SD) being calculated. For categorical variables, the chi-square and Fisher's exact tests were used. Student's t-test was used for quantitative variables, which presented normal distribution.

A logistic regression analysis was performed to build three different univariable prediction models based on different cut-offs (II, III, and IV) of the MMT and to build their ROC curves.

For the sample size estimation, we assumed MMT sensitivity to be 70% at the threshold of class III.¹⁴ A target sensitivity of 90% was then assumed to improve test performance at a different threshold. A sample of 31 positive outcomes would be needed to catch the sensitivity of 90% for the MMT with the new threshold for a *p*-value of 0.048 and a power of 0.807.²¹ Considering 20% of possible missing data, we aimed to get 37 difficult laryngoscopies. A total of 578 patients would be necessary considering 6.4% of frequency of difficult laryngoscopies.²²

Results

From the initial 573 eligible patients, 570 were submitted to statistical analysis, with 3 being excluded because of missing data on Cormack and Lehane's classification (Fig. 1). Difficult laryngoscopy was presented by 36 patients (6.32%). Modified Mallampati classification was reported as follows: 214 with class I (37.54%), 191 with class II (33.51%), 113 with class III (19.82%), and 51 with class IV (8.95%). Table 1 summarizes the descriptive analysis of demographic data and ASA physical status.

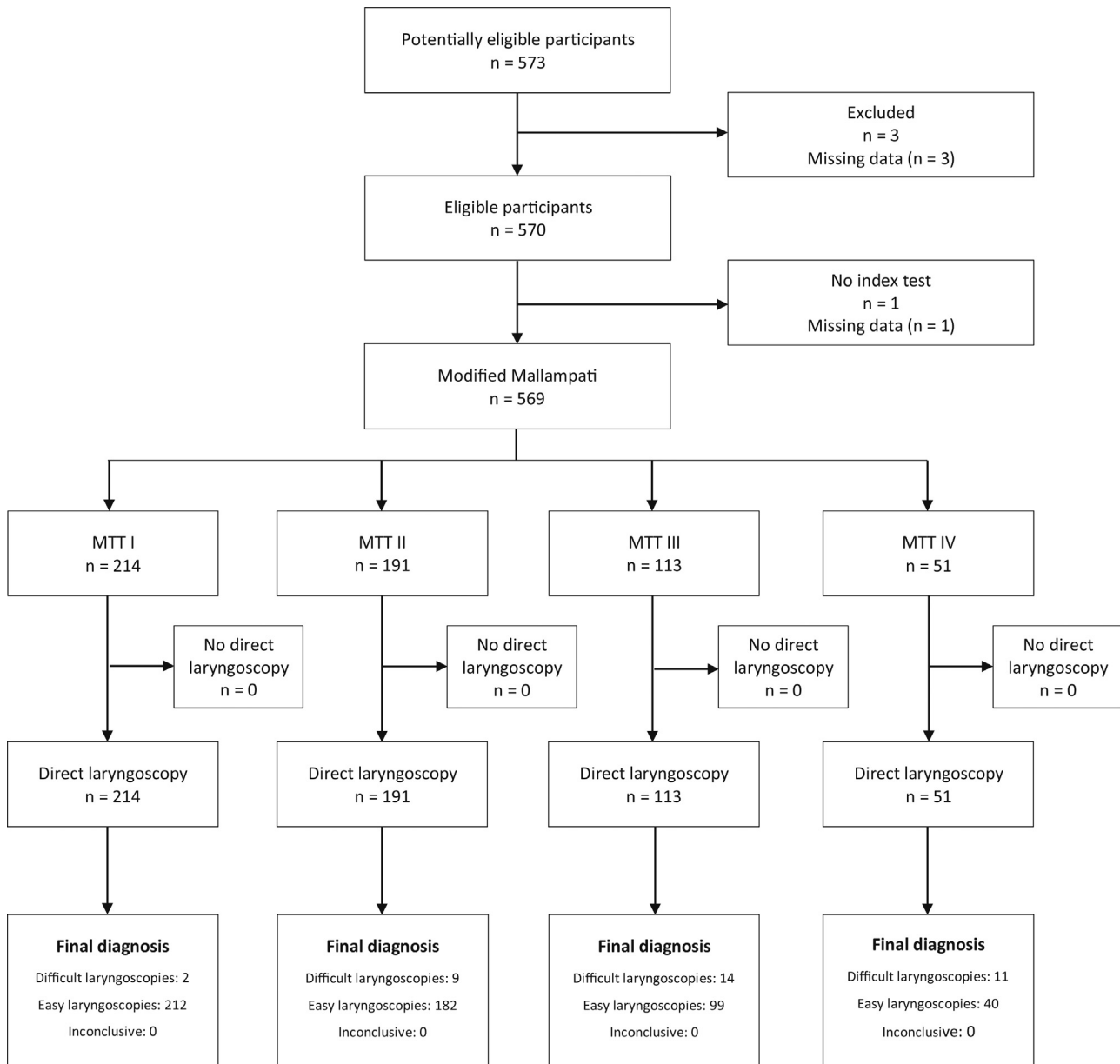


Figure 1 STARD flow chart.

Table 2 shows the analyses based on chi-square and Fisher's exact tests. These analyses found association between difficult laryngoscopy and sex, ASA physical status, and MMT. The male gender was indicative of greater odds of difficult laryngoscopy (OR: 3.119; 95% CI:1.476–6.866; $p = 0.001$). The cut-off for ASA physical status was chosen based on Youden Index, being $ASA \geq 2$ associated with difficult laryngoscopy (OR: 2.71; 95% CI:1.21–6.69; $p = 0.014$). For the MMT, using the class III as the cut-off point for difficult airways showed the highest balanced accuracy (71.68%; 95% CI: 63.83–79.54%). However, the class II was the cut-off point related to the highest increase in odds of facing a difficult laryngoscopy (OR: 11.2; 95% CI:2.82–97.23; $p = 0.000$) as well as to the highest sensitivity (94.44%).

The Student's t -test demonstrated difficult laryngoscopies to be associated with age ($p = 0.032$), weight ($p = 0.000$), and height ($p = 0.000$).

The statistical analysis made to evaluate the predictive values of MMT at different cut-off points is summarized in Table 3. Also, a ROC curve was constructed to compare the predictive performance of MMT at different cut-offs (Fig. 2).

Three predictive logistic regression models for difficult laryngoscopy were constructed including the statistically significant variables at the bivariate analyses (Table 4). Each model accounted for a different MMT threshold.

Discussion

The main result of the present study is the considerable highest sensitivity (94.44%) demonstrated by the test 1 as compared to the tests 2 (69.44%) and 3 (30.56%). Therefore, using Mallampati class II as the cut-off point for possible difficult laryngoscopies performs better as a screening test

Table 1 Demographic data and ASA physical status. Values are mean (SD) or n (%).

Characteristics	Total n = 570	Laryngoscopy:		p-value
		easy and difficult n = 534	n = 36	
Age (years)	43.5 (17.4)	43.1 (17.2)	50 (18.2)	0.033
Height (cm)	161.2 (12.5)	160.7 (12.5)	168.6 (9.9)	0.000
Weight (kg)	70.9 (21.2)	70 (20.8)	84.1 (22.4)	0.001
BMI (kg. m ⁻²)				
<30	437 (76.6)	413 (77.3)	24 (66.7)	0.206
≥30 (obese)	133 (23.4)	121 (22.7)	12 (33.3)	
Sex				
Female	354 (62.1)	341 (63.9)	13 (36.1)	0.001
Male	216 (37.9)	193 (36.1)	23 (63.9)	
ASA				
I	263 (46.1)	254 (47.6)	9 (25)	0.014
II	236 (41.4)	218 (40.8)	18 (50)	
III	69 (12.1)	60 (11.2)	9 (25)	
IV	2 (0.4)	2 (0.4)	–	

SD, standard deviation; ASA, physical status classification according to the American Society of Anesthesiologists system; BMI, body mass index; < Less; ≥ greater or equal.

Table 2 Distribution of the patients according to preoperative predictors in the bivariate analysis as a function of whether laryngoscopy was easy or difficult.

Variables		Laryngoscopy		Bivariate analysis	
		Easy	Difficult	Odds ratio	Chi-square
		N	N	Fisher (95%CI)	p-value
Sex	Female	341	13	3.119 (1.476–6.866)	0.001
	Male	193	23		
Obesity	Non-obese	413	24	1.704 (0.753–3.668)	0.206
	Obese	121	12		
ASA	I	254	9	2.71 (1.21–6.69)	0.014
	≥ II	280	27		
Test 1 (Mallampati)	Easy (I)	212	2	11.20 (2.82–97.23)	0.000
	Difficult (II, III, IV)	321	34		
Test 2 (Mallampati)	Easy (I, II)	394	11	6.42 (2.95–14.85)	0.000
	Difficult (III, IV)	139	25		
Test 3 (Mallampati)	Easy (I, II, III)	493	25	5.40 (2.23–12.39)	0.000
	Difficult (IV)	40	11		

Tests 1, 2, and 3 are modified Mallampati tests as follows: test 1, easy (Mallampati I) or difficult (II, III, and IV); test 2, easy (I and II) or difficult (III and IV); test 3, easy (I, II, and III) or difficult (IV). Odds ratios are presented for bottom classes as related to top classes (e.g., male has 3.119 more chances of presenting difficult laryngoscopies as compared to female).

Table 3 Predictive values for three different approaches of modified Mallampati test to predict the occurrence of difficult laryngoscopy.

Model	Sens	PPV	Spec	NPV	Bal Acc (95% CI)	Acc (95% CI)
Test 1	94.44%	9.57%	39.77%	99.06%	67.11% (62.78–71.44%)	43.23% (39.12–47.42%)
Test 2	69.44%	15.24%	73.92%	97.28	71.68% (63.83–79.54)	73.64% (69.81–77.22%)
Test 3	30.56%	21.57%	92.49%	95.17%	61.53% (53.81–69.24%)	88.58% (85.67–91.07%)

TP, true positive; TN, true negative; FP, false positive; FN, false negative; Sens, sensitivity; PPV, positive predictive value; Spec, specificity; NPV, negative predictive value; Bal Acc, balanced accuracy; Acc, overall accuracy; 95% CI, confident interval 95%; test 1: Mallampati II, III, and IV as difficult; test 2: Mallampati III and IV as difficult; test 3: Mallampati IV as difficult.

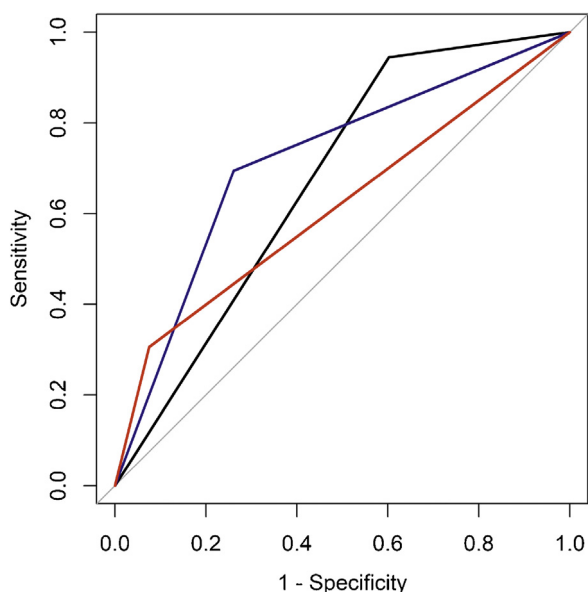


Figure 2 ROC curves for different approaches of modified Mallampati test. Black line: Mallampati II, III, and IV as difficult (AUC = 67.11%); blue line: Mallampati III and IV as difficult (AUC = 71.68%); red line: Mallampati IV as difficult (AUC = 61.53%).

than the other thresholds. Seemingly endorsing test 1 as the best screening test is its highest odds ratio (OR: 11.20) demonstrating the highest increase in chances of facing a difficult laryngoscopy when the test is indicative of doing so (test 2 OR: 6.42; test 3 OR: 5.40).

The MMT has been used in a bid to correctly segregate easy and difficult airways. As trying to do so, its performance has been demonstrating to be only of poor value.^{6,10,14} A recent meta-analysis summarized the frequency of difficult laryngoscopies as 11% (95% CI: 6–19%). The same study reported MMT sensitivity as 53% (95% CI: 47–59%) and specificity as 80% (95% CI: 74–85%).¹⁴ It makes around half of difficult laryngoscopies unanticipated and hence likely to be improperly approached. Besides, such sensitivity and

specificity values lead to an overall accuracy of 77.3%. In turn, setting every patient as easy would present a considerable higher overall accuracy of 89% (100% [all patients]–11% [frequency of difficult laryngoscopies]). Additionally, these predictive values are linked to a balanced accuracy of 66.5%, also demonstrating a poor ability of the MMT to efficiently segregate easy and difficult laryngoscopies. In brief, as advocated by Yentis, to correctly classify the largest number of patients, assigning all patients as easy would do better than MMT, while the attempt to discriminate difficult laryngoscopies through the use of the conventional MMT still remains a practice of low sensitivity.⁸

On the other hand, it seems clear the focus of preoperative upper airway assessment should be on hitting the difficult airways. These are the patients who would suffer the worst damage if unanticipated. To achieve this target, a screening approach is more useful than an attempt to correctly classify the largest number of patients.^{9,15,23} This way, highest sensitivity and odds ratio would fit better than highest overall and balanced accuracy. It might be the focus of the MMT and, accordingly, in the present study, the best screening performance for the Mallampati classification was achieved with the class II as the cut-off – test 1.

Additionally, the test 1 presented a balanced accuracy (Bal Acc = 67.11%; 95% CI: 62.78–71.44%) comparable to the best one, of the test 2 (Bal Acc = 71.68%; 95% CI: 63.83–79.54%). This is achieved by the compensation between the high sensitivity and low specificity of the test 1 and the high specificity and low sensitivity of the test 2. In other words, when changing the threshold from Mallampati class III to class II, the anesthesiologist apparently transfers comparable proportion of correct classification from easy laryngoscopies to difficult ones. It seems to make test 1 more clinically useful since the focus is supposed to be on difficult airways.

It's worth noticing that when taking the Mallampati classification for a screening approach by changing the threshold from class III to class II, the test is less likely to correctly hit a difficult laryngoscopy when predicting so. We should keep in mind that the purpose shifts as well – we are no longer interested in getting all predictions right. On top of this,

Table 4 Summary of the three multivariable models including Mallampati test at different thresholds for prediction of difficult laryngoscopies. Model I includes test I (ease: Mallampati I; difficulty: Mallampati II-IV). Model II includes test II (ease: Mallampati I-II; difficulty: Mallampati III-IV). Model III includes test III (ease: Mallampati I-III; difficulty: Mallampati IV). Rows with $p < 0.05$ are highlighted.

Characteristic	Model I			Model II			Model III		
	OR	95% CI	$p > z $	OR	95% CI	$p > z $	OR	95% CI	$p > z $
(Intercept)	0.000	0.000–0.000	0.000	0.000	0.000–0.000	0.000	0.000	0.000–0.000	0.000
Sex (male)	1.024	0.378–2.776	0.963	1.176	0.431–3.209	0.752	1.278	0.472–3.459	0.629
Mallampati	10.065	2.33–43.420	0.002	5.026	2.363–10.688	0.000	4.141	1.773–9.670	0.001
Age	1.015	0.990–1.042	0.244	1.014	0.986–1.041	0.304	1.017	0.992–1.044	0.186
Weight	1.020	1.001–1.039	0.035	1.020	1.002–1.039	0.033	1.020	1.002–1.039	0.030
Height	1.063	1.009–1.119	0.021	1.056	1.002–1.114	0.042	1.053	1.001–1.108	0.047
ASA (≥ 2)	1.23	0.677–2.231	0.499	1.322	0.725–2.411	0.363	1.322	0.720–2.429	0.368

ASA, American Society of Anesthesiologists physical status; OR, odds ratio; 95% CI, 95% confidence interval.

it's worth highlighting that test accuracy is dependent on how frequent an outcome is and that the positive predictive value is specially impaired by the relatively low frequency of difficult laryngoscopies –, not bringing further concern to the screening performance of the test.

ASA physical status was also associated with difficult laryngoscopies in the bivariate analysis, showing an odds 2.7 times greater of facing difficult laryngoscopies when managing patients with ASA II or higher. It may reflect anatomic alterations caused by systemic diseases, although this association found should be interpreted with caution, also because it was lost in the multivariable analysis.

Some limitations must be taken into account. The relatively small number of difficult laryngoscopies impairs more conclusive results. The interobserver variability of the Mallampati classification as well as of the Cormack and Lehane graduation system may have also skewed the conclusions. Additionally, the lack of a standard protocol of anesthetic induction and the lack of a neuromuscular-blockade-depth measurement may have interfered the degree of difficulty during the airway manipulations.

Conclusions

In conclusion, the results of the present study suggest the Mallampati class II as the more clinically useful threshold for airway prediction. At this cut-off, anesthesiologists would be able to anticipate further difficult laryngoscopies making airway manipulation a safest intervention. This way, we recommend airway managers to be more careful with their approaching strategy for airway manipulation from MMT class II.

Funding

Departmental funding only.

Conflicts of interest

The authors declare no conflicts of interest.

References

- Cook TM, Woodall N, Harper J, et al. Fourth National Audit Project. Major complications of airway management in the UK: results of the Fourth National Audit Project of the Royal College of Anaesthetists and the Difficult Airway Society. Part 2: intensive care and emergency departments. *Br J Anaesth*. 2011;106:632–42.
- Cook TM, Woodall N, Frerk C. Fourth National Audit Project. Major complications of airway management in the UK: results of the Fourth National Audit Project of the Royal College of Anaesthetists and the Difficult Airway Society. Part 1: anaesthesia. *Br J Anaesth*. 2011;106:617–31.
- Tikka T, Hilmi OJ. Upper airway tract complications of endotracheal intubation. *Br J Hosp Med*. 2019;80:441–7.
- Umobong EU, Mayo PH. Critical care airway management. *Crit Care Clin*. 2018;34:313–24.
- Higgs A, McGrath BA, Goddard C, et al. Guidelines for the management of tracheal intubation in critically ill adults. *Br J Anaesth*. 2018;120:323–52.
- Vannucci A, Cavallone LF. Bedside predictors of difficult intubation: a systematic review. *Minerva Anestesiologica*. 2016;82:69–83.
- Baker P. Assessment before airway management. *Anesthesiol Clin*. 2015;332:257–78.
- Yentis SM. Predicting difficult intubation—worthwhile exercise or pointless ritual? *Anaesthesia*. 2002;572:105–9.
- Pandit JJ, Heidegger T. Putting the “point” back into the ritual: a binary approach to difficult airway prediction. *Anaesthesia*. 2017;723:283–8.
- Shiga T, Wajima Z, Inoue T, et al. Predicting difficult intubation in apparently normal patients: a meta-analysis of bedside screening test performance. *Anesthesiology*. 2005;103:429–37.
- Nørskov AK, Rosenstock CV, Wetterslev J, et al. Diagnostic accuracy of anaesthesiologists' prediction of difficult airway management in daily clinical practice: a cohort study of 188 064 patients registered in the Danish Anaesthesia Database. *Anaesthesia*. 2015;70:272–81.
- 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.
- Samssoon GLT, Young JRB. Difficult tracheal intubation: a retrospective study. *Anaesthesia*. 1987;42:487–90.
- Roth D, Pace NL, Lee A, et al. Bedside tests for predicting difficult airways: an abridged Cochrane diagnostic test accuracy systematic review. *Anaesthesia*. 2019;74:915–28.
- Teos WH, Kristensen MS. Prediction in airway management: what is worthwhile, what is a waste of time and what about the future? *Br J Anaesth*. 2016;117:1–3.
- Cook TM. A new practical classification of laryngeal view. *Anaesthesia*. 2000;55:274–9.
- R Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2019. Available at: <https://www.R-project.org/>. [Accessed 12 August 2019].
- Kuhn M. caret: Classification and Regression Training. R package version 6.0-85; 2020. Available at: <https://CRAN.R-project.org/package=caret>. [Accessed 25 February 2020].
- Sing T, Sander O, Beerenwinkel N, et al. ROCR: visualizing classifier performance in R. *Bioinformatics*. 2005;21:3940–1.
- Lopez-Raton M, Rodriguez-Alvarez MX, Suarez CC, et al. Optimal cutpoints: an R package for selecting optimal cutpoints in diagnostic tests. *J Stat Software*. 2014;61:1–36.
- Robin X, Turck N, Hainard A, et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics*. 2011;12:77.
- Bujang MA, Adnan TH. Requirements for minimum sample size for sensitivity and specificity analysis. *J Clin Diagn Res*. 2016;10:YE01–6.
- de Carvalho CC, da Silva DM, de Carvalho Junior AD, et al. Pre-operative voice evaluation as a hypothetical predictor of difficult laryngoscopy. *Anaesthesia*. 2019;74:1147–52.