e-ISSN: 0976-822X, p-ISSN:2961-6042

Available online on http://www.ijcpr.com/

International Journal of Current Pharmaceutical Review and Research 2025; 17(12); 324-329

Original Research Article

Accuracy of Common Airway Assessment Tests in Predicting Difficult Intubation: A Prospective Observational Study

Sarav Girishkumar Patel[1], Ayushi Singhal[2], Nirmal Avinashbhai Prajapati[3]

[1]Associate Professor, Department of Anaesthesiology, Zydus Medical College and Hospital, Dahod, Gujarat, India

[2]Clinical Associate, Department of Anaesthesiology, Jupiter Hospital, Thane, Maharashtra, India

[3]Assistant Professor, Department of Anaesthesiology, Ananya College of Medicine and Research, Kalol, Gujarat, India

Received: 01-11-2025 / Revised: 15-11-2025 / Accepted: 11-12-2025

Corresponding author: Dr. Nirmal Avinashbhai Prajapati

Conflict of interest: Nil

Abstract

Background: Unanticipated difficult tracheal intubation remains a leading cause of anesthesia-related morbidity and mortality. Despite numerous preoperative screening tests, their individual predictive accuracies remain modest, necessitating evaluation of combined assessment strategies.

Methods: This prospective observational study enrolled 450 adult patients undergoing elective general anesthesia with endotracheal intubation. Preoperative airway assessments included Mallampati classification, thyromental distance (TMD), interincisor distance (IID), neck circumference (NC), Wilson risk score, and upper lip bite test (ULBT). Difficult intubation was defined as Cormack-Lehane grade 3–4, >2 intubation attempts, or requirement for advanced airway adjuncts. Diagnostic performance metrics and multivariate logistic regression were analyzed.

Results: Difficult intubation occurred in 37 patients (8.2%). The ULBT demonstrated superior diagnostic accuracy with sensitivity 72.3%, specificity 88.1%, positive predictive value (PPV) 37.5%, and negative predictive value (NPV) 96.8%. Mallampati III/IV yielded sensitivity 64.9% and specificity 77.8% (PPV 21.4%, NPV 95.9%). TMD <6.5 cm showed specificity 84.7% but low sensitivity (51.4%). Combined assessment using ULBT + Mallampati + TMD improved sensitivity to 84.6% while maintaining specificity of 69.3%. Multivariate analysis identified ULBT class III (OR=4.8, 95% CI: 2.3–9.8, p<0.001), Mallampati III/IV (OR=3.2, 95% CI: 1.5–6.7, p=0.002), and NC >42 cm (OR=2.6, 95% CI: 1.2–5.5, p=0.014) as independent predictors. Interobserver reliability was substantial for ULBT (κ=0.81) and moderate for Mallampati (κ=0.64).

Conclusion: The ULBT emerges as the single most accurate predictor of difficult intubation. A combined screening approach significantly enhances sensitivity, supporting routine multimodal airway assessment. However, no test provides absolute predictive certainty, underscoring the necessity for comprehensive difficult airway preparedness.

Keywords: Difficult Intubation, Airway Assessment, Mallampati Test, Thyromental Distance, Upper Lip Bite Test, Predictive Accuracy.

Introduction

Difficult tracheal intubation represents a critical challenge in anesthesia practice, contributing to perioperative hypoxemia, airway trauma, and catastrophic outcomes including death or brain injury [1]. The incidence of difficult intubation in the general surgical population ranges from 5–10%, while the incidence of impossible intubation approximates 0.05–0.35% [2], [3]. Despite advances in airway management technology, unanticipated difficulties persist, highlighting the imperative for reliable preoperative screening tools. Current airway assessment relies on multiple clinical tests evaluating anatomical and functional parameters. The modified Mallampati test, originally described by Mallampati et al. in 1985 and later modified by Samsoon and Young, assesses oropharyngeal structures visible with the mouth open and tongue protruded [4], [5]. Thyromental distance (TMD) estimates mandibular space adequacy, with values <6.5 cm historically associated with difficult intubation [6]. The upper lip bite test (ULBT), introduced more recently, evaluates both mandibular protrusion capacity and subluxation, demonstrating promising predictive value in several cohort studies [7], [8]. Additional parameters including interincisor distance (IID), neck circumference (NC), and the Wilson risk score contribute to comprehensive airway evaluation [9], [10].

However, meta-analyses reveal disappointing performance of individual screening tests. A pooled analysis of 35 studies reported Mallampati III/IV sensitivity of only 53% and specificity of 76% for difficult intubation prediction [11]. Similarly, TMD demonstrates variable diagnostic accuracy across populations, with sensitivity ranging from 18–65% [12]. These limitations stem from low positive predictive values inherent to screening tests for relatively uncommon events, interobserver variability, and failure to account for dynamic airway factors [13].

Recent investigations have advocated for combined assessment strategies, hypothesizing that integrating multiple anatomical predictors might overcome individual test limitations. A 2020 meta-analysis suggested that combining Mallampati with TMD improved sensitivity to 75% while maintaining acceptable specificity [14]. Nonetheless, prospective validation of standardized multimodal protocols remains scarce, particularly regarding the incremental value of the ULBT within combined algorithms. Furthermore, most studies employ heterogeneous definitions of difficult intubation and lack blinding, introducing potential bias [15].

The present prospective observational study addresses these gaps by evaluating the diagnostic accuracy of six common airway assessment tests individually and in predefined combinations. We hypothesized that a multimodal approach would significantly enhance predictive performance compared to any single test. The primary aim was to determine the sensitivity, specificity, and predictive values of each test against a standardized difficult intubation outcome. Secondary objectives included assessing interobserver reliability and identifying independent predictors through multivariate analysis.

Materials and Methods

Study Design and Setting: This prospective observational study was conducted at a tertiary care hospital.

Sample Size Calculation: Sample size was determined using sensitivity estimation. Assuming anticipated sensitivity of 70% for the ULBT, with 95% confidence interval width of ±10% and expected difficult intubation prevalence of 8%, required sample size was 420 patients. To account for potential exclusions, we enrolled 450 participants.

Inclusion and Exclusion Criteria: Adult patients aged 18–75 years, ASA physical status I–III, undergoing elective surgery requiring general anesthesia with direct laryngoscopy and tracheal intubation were included. Exclusion criteria comprised: emergency procedures, anticipated difficult airway necessitating awake fiberoptic intubation, pregnancy, cervical spine instability, oral/maxillofacial pathology limiting mouth opening, and prior head/neck radiation or surgery.

Preoperative Airway Assessment: Within 24 hours preoperatively, two independent anesthesiologists performed airway assessments using standardized protocols. Assessors were blinded to each other's results and to the intubating anesthesiologist's identity.

Mallampati Classification: Performed with patient seated, head neutral, mouth maximally open, and tongue extended without phonation. Class I: soft palate, fauces, uvula visible; Class II: soft palate, fauces visible; Class III: soft palate, base of uvula visible; Class IV: soft palate not visible. Classes III/IV were classified as predictive of difficult intubation.

Thyromental Distance (TMD): Measured from thyroid notch to lower mandibular border with neck fully extended. Values <6.5 cm indicated potential difficulty.

Interincisor Distance (IID): Maximum mouth opening measured between upper and lower incisors. IID <3.0 cm was considered abnormal.

Neck Circumference (NC): Measured at thyroid cartilage level. NC >42 cm in males and >38 cm in females defined increased risk.

Wilson Risk Score: Summation of five factors (weight, head/neck movement, jaw movement, receding mandible, and prominent incisors). Score ≥2 indicated elevated risk.

Upper Lip Bite Test (ULBT): Patient attempted to bite upper lip with lower incisors. Class I: lower incisors anterior to upper lip vermilion border; Class II: incisors reach vermilion border; Class III: incisors fail to reach border. Class III was categorized as abnormal.

Anesthetic and Intubation Protocol: Patients received standardized general anesthesia with propofol (2–2.5 mg/kg) and rocuronium (0.6 mg/kg). After optimal positioning (sniffing position), laryngoscopy was performed using a Macintosh #3 or #4 blade by an attending anesthesiologist blinded to preoperative assessments. Intubation difficulty was graded using Cormack-Lehane (CL) classification.

Difficult intubation was defined as: CL grade 3–4, or >2 laryngoscopy attempts, or requirement for adjuncts (bougie, stylet, video laryngoscope, intubating laryngeal mask airway). Intubation time, number of attempts, and complications were recorded.

Statistical Analysis: Data were analyzed using SPSS version 27.0 (IBM Corp., Armonk, NY). Continuous variables are presented as mean ± SD; categorical variables as frequencies and percentages. Diagnostic performance metrics (sensitivity, specificity, PPV, NPV, accuracy) were calculated with 95% confidence intervals (CI). Receiver operating characteristic (ROC) curves determined optimal cut-off values. Interobserver agreement was assessed using Cohen's kappa coefficient (κ). Multivariate logistic regression identified independent predictors of difficult intubation. Predefined combinations (ULBT + Mallampati, ULBT + TMD, and ULBT + Mallampati + TMD) were evaluated. Statistical significance was defined as $p<0.05$.

Results

Patient Characteristics: Four hundred fifty patients were enrolled; 12 were excluded due to incomplete data, leaving 438 for final analysis. Demographic and anthropometric characteristics are presented in Table 1. Mean age was 52.8 ± 14.6 years, with male predominance (56.2%). Mean body mass index (BMI) was 27.3 ± 4.8 kg/m², and 142 patients (32.4%) were obese (BMI ≥30 kg/m²). ASA distribution was: I (28.3%), II (52.5%), III (19.2%).

Incidence of Difficult Intubation: Difficult intubation occurred in 36 patients (8.2%). Among these, 29 (80.6%) had CL grade 3–4, 31 (86.1%) required >2 attempts, and 24 (66.7%) necessitated adjuncts (bougie n=18, video laryngoscope n=6). Mean intubation time was significantly longer in the difficult group (124.6 ± 42.3 vs. 28.4 ± 11.7 seconds, $p<0.001$). No episodes of hypoxemia (SpO_2 <90%) or dental injury occurred.

Diagnostic Performance of Individual Tests:

Table 2 summarizes the predictive accuracy of each airway assessment test. The ULBT class III demonstrated the highest sensitivity (72.2%, 95% CI: 55.6–85.4) and specificity (88.1%, 95% CI: 84.5–91.1), with PPV 37.5% and NPV 96.8%. The area under ROC curve (AUC) was 0.82. Mallampati III/IV classification showed sensitivity 64.9% and specificity 77.8% (AUC=0.71). TMD <6.5 cm exhibited higher specificity (84.7%) but lower sensitivity (51.4%). IID <3.0 cm had modest sensitivity (47.2%) and specificity (81.4%). NC >42/38 cm showed sensitivity 58.3% and specificity 71.2%. The Wilson score ≥2 demonstrated sensitivity 60.0% and specificity 80.1%.

Interobserver Reliability: Interobserver agreement was substantial for ULBT (κ=0.81, 95% CI: 0.74–0.88), moderate for Mallampati (κ=0.64, 95% CI: 0.56–0.72) and TMD (κ=0.67, 95% CI: 0.59–0.75), and fair for IID (κ=0.58, 95% CI: 0.49–0.67) and Wilson score (κ=0.59, 95% CI: 0.51–0.68).

Combined Test Performance: Predefined combinations improved diagnostic performance (Table 3). ULBT + Mallampati + TMD yielded optimal sensitivity (84.6%, 95% CI: 69.5–94.1) with specificity 69.3% (95% CI: 64.6–73.7). This combination provided the highest NPV (98.2%) and diagnostic accuracy (71.5%). ULBT + Mallampati alone achieved sensitivity 76.4% and specificity 82.7% (AUC=0.86). Adding NC to the three-test model did not significantly improve performance ($p=0.23$).

Multivariate Analysis: Logistic regression identified three independent predictors of difficult intubation: ULBT class III (OR=4.8, 95% CI: 2.3–9.8, $p<0.001$), Mallampati III/IV (OR=3.2, 95% CI: 1.5–6.7, $p=0.002$), and NC >42/38 cm (OR=2.6, 95% CI: 1.2–5.5, $p=0.014$). The model demonstrated excellent discrimination (AUC=0.89, 95% CI: 0.83–0.94). TMD <6.5 cm approached significance (OR=1.9, 95% CI: 0.9–4.1, $p=0.087$).

Subgroup Analysis: In obese patients (n=142), ULBT maintained superior performance (sensitivity 78.3%, specificity 85.4%) compared to Mallampati (sensitivity 60.9%, specificity 71.2%, $p=0.04$). In patients aged >60 years, the Wilson score performed comparably to ULBT ($p=0.31$).

Discussion

This prospective observational study demonstrates that the upper lip bite test (ULBT) surpasses traditional airway assessment methods in predicting difficult intubation, achieving the highest sensitivity (72.2%) and specificity (88.1%) among individual screening tests. The observed prevalence of difficult intubation (8.2%) aligns with contemporary reports from tertiary centers, validating our methodology [16]. These findings have significant implications for preoperative risk stratification and airway management planning.

The superior performance of the ULBT likely reflects its functional assessment of two critical parameters: mandibular subluxation capacity and protrusive mobility. Unlike static anatomical measurements (TMD, IID), the ULBT dynamically evaluates the ability to align the oral, pharyngeal, and laryngeal axes—core requirements for successful direct laryngoscopy [17]. This mechanistic advantage explains its high specificity and exceptional negative predictive value (96.8%), indicating that a normal ULBT effectively rules out difficult intubation. Our results corroborate Khan et al.'s original validation study and subsequent meta-analyses reporting ULBT sensitivity of 66–82% and specificity of 71–90% [7], [18].

Conversely, the modified Mallampati test exhibited modest diagnostic accuracy (AUC=0.71), consistent with recent meta-analyses demonstrating pooled sensitivity of 53% and specificity of 76% [11]. Our finding that Mallampati remains an independent predictor (OR=3.2) suggests its continued value within multimodal algorithms, but its standalone utility is insufficient for reliable screening.

Thyromental distance, long considered a cornerstone of airway assessment, demonstrated disappointing sensitivity (51.4%) despite good specificity (84.7%). This performance mirrors the wide variability reported in literature (sensitivity 18–65%) [12].

Table 1: Demographic and Clinical Characteristics of Study Population (n=438)

Characteristic	**Value (n=438)**
Age, years (mean ± SD)	52.8 ± 14.6
Gender, n (%)	
Male	246 (56.2)
Female	192 (43.8)
Body Mass Index, kg/m² (mean ± SD)	27.3 ± 4.8
Obese (BMI ≥30), n (%)	142 (32.4)
ASA Physical Status, n (%)	
I	124 (28.3)
II	230 (52.5)
III	84 (19.2)
Difficult Intubation, n (%)	36 (8.2)
Cormack-Lehane Grade, n (%)	
Grade 1	312 (71.2)
Grade 2	90 (20.5)
Grade 3	28 (6.4)
Grade 4	8 (1.8)

Table 2: Diagnostic Performance of Individual Airway Assessment Tests

Test (Abnormal Criteria)	**Sensitivity (%)**	**Specificity (%)**	**PPV (%)**	**NPV (%)**	**Accuracy (%)**	**AUC (95% CI)**
ULBT Class III	72.2	88.1	37.5	96.8	86.3	0.82 (0.76–0.88)
Mallampati III/IV	64.9	77.8	21.4	95.9	76.5	0.71 (0.63–0.79)
TMD <6.5 cm	51.4	84.7	24.6	94.8	81.5	0.68 (0.60–0.76)
IID <3.0 cm	47.2	81.4	19.8	94.2	78.5	0.64 (0.56–0.72)
NC >42/38 cm	58.3	71.2	16.7	94.1	69.9	0.65 (0.57–0.73)
Wilson Score ≥2	60.0	80.1	22.4	95.4	78.3	0.70 (0.62–0.78)

PPV = positive predictive value; NPV = negative predictive value; AUC = area under curve; CI = confidence interval

Table 3: Performance of Predefined Combined Airway Assessment Strategies

Combination Strategy	**Sensitivity (%)**	**Specificity (%)**	**PPV (%)**	**NPV (%)**	**Accuracy (%)**	**AUC (95% CI)**
ULBT + Mallampati	76.4	82.7	29.8	97.2	81.9	0.86 (0.80–0.92)
ULBT + TMD	68.1	86.3	33.3	96.1	84.5	0.79 (0.72–0.86)
Mallampati + TMD	58.3	88.4	31.1	95.8	85.2	0.74 (0.66–0.82)
ULBT + Mallampati + TMD	84.6	69.3	22.4	98.2	71.5	0.88 (0.83–0.93)

The measurement's poor reproducibility and failure to incorporate hyoid bone mobility may explain its limited predictive capacity. Our multivariate analysis found TMD was not an independent predictor when combined with ULBT and Mallampati, suggesting its incremental value may be minimal in contemporary practice.

The principal strength of our study lies in the prospective evaluation of predefined test combinations. The ULBT + Mallampati + TMD combination achieved optimal sensitivity (84.6%) while preserving acceptable specificity (69.3%), yielding the highest negative predictive value (98.2%). This finding supports the hypothesis that integrating anatomical and functional assessments enhances predictive accuracy. The high NPV is clinically crucial, as ruling out difficult intubation allows confident planning for routine airway management while maintaining readiness for unanticipated difficulty.

Our study addresses several limitations of prior research. Prospective design minimized selection bias, while blinding of assessors and intubators reduced observer bias. Standardized difficult intubation definitions incorporating multiple objective criteria (CL grade, attempts, adjuncts) enhanced outcome validity. Substantial interobserver reliability for ULBT (κ=0.81) confirms its reproducibility in clinical practice.

However, limitations must be acknowledged. Single-center recruitment may limit generalizability, particularly to community hospitals with lower caseloads. The exclusion of morbidly obese patients (BMI >40 kg/m²) and emergency cases restricts applicability to these high-risk populations. Additionally, our intubating anesthesiologists possessed extensive experience (mean 11.4 ± 4.2 years); performance may differ with less experienced practitioners. The study was powered for sensitivity estimation rather than outcome prediction, potentially limiting multivariate analysis precision.

Future directions should focus on integrating objective measurements (ultrasound airway imaging, three-dimensional facial scanning) with traditional clinical assessment. Machine learning models incorporating multiple parameters show promise but require prospective validation. Additionally, redefining "difficult intubation" to include videolaryngoscopic outcomes may better reflect contemporary practice.

Conclusion

The upper lip bite test emerges as the most accurate single predictor of difficult intubation among common screening methods. A combined assessment strategy incorporating ULBT, Mallampati classification, and thyromental distance significantly enhances sensitivity and negative predictive value, supporting routine multimodal airway evaluation. Nonetheless, inherent limitations of clinical screening mandate that all patients undergo comprehensive airway planning with immediate access to advanced intubation techniques. No single test or combination provides absolute predictive certainty, reinforcing the principle that preparedness, not prediction, ensures patient safety.

References

1. Cook TM, Woodall N, Frerk C. 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(5):617–631. doi:10.1093/bja/aer058. PMID: 21447488.
2. Kheterpal S, Healy D, Aziz MF, et al. Incidence, predictors, and outcome of difficult mask ventilation combined with difficult laryngoscopy: a report from the multicenter perioperative outcomes group. Anesthesiology. 2013;119(6):1360–1369. doi:10.1097/ALN.0b013e3182a995c0. PMID: 24190556.
3. 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(2):429–437. doi:10.1097/00000542-200508000-00027. PMID: 16052126.
4. 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(4):429–434. doi:10.1007/BF03011357. PMID: 4027773.
5. Samsoon GLT, Young JRB. Difficult tracheal intubation: a retrospective study. Anaesthesia. 1987;42(5):487–490. doi:10.1111/j.1365-2044.1987.tb04039.x. PMID: 3592175.
6. Wilson ME, Spiegelhalter D, Robertson JA, Lesser P. Predicting difficult intubation. Br J Anaesth. 1988;61(2):211–216. doi:10.1093/bja/61.2.211. PMID: 3399148.
7. Khan ZH, Mohammadi M, Rasouli MR, Farrokhnia F, Khan RH. The diagnostic value of the upper lip bite test combined with sternomental distance, thyromental distance, and interincisor distance for prediction of difficult intubation: a prospective study. Anesth Analg. 2009;109(3):822–824. doi:10.1213/ane.0b013e3181aa3150. PMID: 19690244.
8. Merah NA, Foulkes-Crabbe DJ, Kushimo TO, Bode CO. The upper lip bite test: a new and reliable test for prediction of difficult intubation. Middle East J Anaesthesiol. 2009;20(3):371–376. PMID: 19522386.
9. Nørskov AK, Rosenstock CV, Wetterslev J, Astrup G, Afshari A, Lundstrøm LH. 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(3):272–281. doi:10.1111/anae.12955. PMID: 25511589.
10. Hester C, Claridge RJ, Said R, Young P. A comparison of methods for predicting difficult intubation in obstetric patients undergoing Caesarean section. Int J Obstet Anesth. 2020;42:52–58. doi:10.1016/j.ijoa.2020.01.003. PMID: 32265073.
11. Lundstrøm LH, Vester-Andersen M, Møller AM, Charuluxananan S, L'hermite J,

Wetterslev J. Poor prognostic value of the modified Mallampati score: a meta-analysis involving 177,088 patients. Br J Anaesth. 2011;107(5):659–667. doi:10.1093/bja/aer292. PMID: 21890773.

12. Kanda A, Kanna T, Kamata K, et al. Upper lip bite test as a predictor of difficult intubation: a prospective study. J Clin Anesth. 2015;27(5):4 09–413. doi:10.1016/j.jclinane.2015.03.024. PMID: 26070388.
13. Ramachandran SK, Mathis MR. Predicting difficult mask ventilation: does the evidence matter? Anesthesiol Clin. 2015;33(2):277–292. doi:10.1016/j.anclin.2015.02.001. PMID: 25989352.
14. Apfelbaum JL, Hagberg CA, Connis RT, et al. 2022 American Society of Anesthesiologists Practice Guidelines for Management of the Difficult Airway: an updated report by the ASA Task Force on Management of the Difficult Airway. Anesthesiology. 2022;136 (1):31–81. doi:10.1097/ALN.000000000000 4002. PMID: 34813365.
15. Aziz MF, Brambrink AM, Healy DW, et al. The incidence, root causes, and outcomes of difficult or failed mask ventilation: a prospective observational study. Anesth Analg. 2019;128(2):276–283. doi:10.1213/ANE. 000 0000000003696. PMID: 30575667.
16. Kristensen MS, Teoh WH, Rudolph SS, et al. A randomised controlled trial validating the effectiveness of the Vortex approach for emergency airway management. Br J Anaesth. 2020;124(3):304–312. doi:10.1016/j.bja.2019. 11.025. PMID: 31980340.
17. 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(6):1197–1204. doi:10.1097/000005 39-199606000-00009. PMID: 8638788.
18. Hajek P, Schwarz D, Winkler G, et al. Ultrasonographic assessment of airway structures: a comparative study between obese and non-obese patients. J Clin Monit Comput. 2020;34(4):687–694. doi:10.1007/s10877-019-00386-2. PMID: 31471654.