Cochrane Library

Cochrane Database of Systematic Reviews

Airway physical examination tests for detection of difficult airway management in apparently normal adult patients (Review)

Roth D, Pace NL, Lee A, Hovhannisyan K, Warenits AM, Arrich J, Herkner H

Roth D, Pace NL, Lee A, Hovhannisyan K, Warenits AM, Arrich J, Herkner H.
Airway physical examination tests for detection of difficult airway management in apparently normal adult patients.
Cochrane Database of Systematic Reviews 2018, Issue 5. Art. No.: CD008874.
DOI: 10.1002/14651858.CD008874.pub2.

www.cochranelibrary.com

WILEY

TABLE OF CONTENTS

[Diagnostic Test Accuracy Review]

Airway physical examination tests for detection of difficult airway management in apparently normal adult patients

Dominik Roth[1], Nathan L Pace[2], Anna Lee[3,4], Karen Hovhannisyan[5], Alexandra-Maria Warenits[1], Jasmin Arrich[1], Harald Herkner[1]

[1]Department of Emergency Medicine, Medical University of Vienna, Vienna, Austria. [2]Department of Anesthesiology, University of Utah, Salt Lake City, UT, USA. [3]Department of Anaesthesia and Intensive Care, The Chinese University of Hong Kong, Shatin, Hong Kong. [4]Hong Kong Branch of The Chinese Cochrane Centre, The Jockey Club School of Public Health and Primary Care, Faculty of Medicine, The Chinese University of Hong Kong, Shatin, Hong Kong. [5]Clinical Health Promotion Centre, Faculty of Medicine, Lund University, Malmö, Sweden

Contact address: Harald Herkner, Department of Emergency Medicine, Medical University of Vienna, Währinger Gürtel 18-20, Vienna, A-1090, Austria. harald.herkner@meduniwien.ac.at.

Editorial group: Cochrane Anaesthesia, Critical and Emergency Care Group.
Publication status and date: New, published in Issue 5, 2018.

Citation: Roth D, Pace NL, Lee A, Hovhannisyan K, Warenits AM, Arrich J, Herkner H. Airway physical examination tests for detection of difficult airway management in apparently normal adult patients. *Cochrane Database of Systematic Reviews* 2018, Issue 5. Art. No.: CD008874. DOI: 10.1002/14651858.CD008874.pub2.

ABSTRACT

Background

The unanticipated difficult airway is a potentially life-threatening event during anaesthesia or acute conditions. An unsuccessfully managed upper airway is associated with serious morbidity and mortality. Several bedside screening tests are used in clinical practice to identify those at high risk of difficult airway. Their accuracy and benefit however, remains unclear.

Objectives

The objective of this review was to characterize and compare the diagnostic accuracy of the Mallampati classification and other commonly used airway examination tests for assessing the physical status of the airway in adult patients with no apparent anatomical airway abnormalities. We performed this individually for each of the four descriptors of the difficult airway: difficult face mask ventilation, difficult laryngoscopy, difficult tracheal intubation, and failed intubation.

Search methods

We searched major electronic databases including CENTRAL, MEDLINE, Embase, ISI Web of Science, CINAHL, as well as regional, subject specific, and dissertation and theses databases from inception to 16 December 2016, without language restrictions. In addition, we searched the Science Citation Index and checked the references of all the relevant studies. We also handsearched selected journals, conference proceedings, and relevant guidelines. We updated this search in March 2018, but we have not yet incorporated these results.

Selection criteria

We considered full-text diagnostic test accuracy studies of any individual index test, or a combination of tests, against a reference standard. Participants were adults without obvious airway abnormalities, who were having laryngoscopy performed with a standard laryngoscope and the trachea intubated with a standard tracheal tube. Index tests included the Mallampati test, modified Mallampati test, Wilson risk score, thyromental distance, sternomental distance, mouth opening test, upper lip bite test, or any combination of these. The target condition was difficult airway, with one of the following reference standards: difficult face mask ventilation, difficult laryngoscopy, difficult tracheal intubation, and failed intubation.

Data collection and analysis

We performed screening and selection of the studies, data extraction and assessment of methodological quality (using QUADAS-2) independently and in duplicate. We designed a Microsoft Access database for data collection and used Review Manager 5 and R for data analysis. For each index test and each reference standard, we assessed sensitivity and specificity. We produced forest plots and summary receiver operating characteristic (ROC) plots to summarize the data. Where possible, we performed meta-analyses to calculate pooled estimates and compare test accuracy indirectly using bivariate models. We investigated heterogeneity and performed sensitivity analyses.

Main results

We included 133 (127 cohort type and 6 case-control) studies involving 844,206 participants. We evaluated a total of seven different prespecified index tests in the 133 studies, as well as 69 non-prespecified, and 32 combinations. For the prespecified index tests, we found six studies for the Mallampati test, 105 for the modified Mallampati test, six for the Wilson risk score, 52 for thyromental distance, 18 for sternomental distance, 34 for the mouth opening test, and 30 for the upper lip bite test. Difficult face mask ventilation was the reference standard in seven studies, difficult laryngoscopy in 92 studies, difficult tracheal intubation in 50 studies, and failed intubation in two studies. Across all studies, we judged the risk of bias to be variable for the different domains; we mostly observed low risk of bias for patient selection, flow and timing, and unclear risk of bias for reference standard and index test. Applicability concerns were generally low for all domains. For difficult laryngoscopy, the summary sensitivity ranged from 0.22 (95% confidence interval (CI) 0.13 to 0.33; mouth opening test) to 0.67 (95% CI 0.45 to 0.83; upper lip bite test) and the summary specificity ranged from 0.80 (95% CI 0.74 to 0.85; modified Mallampati test) to 0.95 (95% CI 0.88 to 0.98; Wilson risk score). The upper lip bite test for diagnosing difficult laryngoscopy provided the highest sensitivity compared to the other tests ($P < 0.001$). For difficult tracheal intubation, summary sensitivity ranged from 0.24 (95% CI 0.12 to 0.43; thyromental distance) to 0.51 (95% CI 0.40 to 0.61; modified Mallampati test) and the summary specificity ranged from 0.87 (95% CI 0.82 to 0.91; modified Mallampati test) to 0.93 (0.87 to 0.96; mouth opening test). The modified Mallampati test had the highest sensitivity for diagnosing difficult tracheal intubation compared to the other tests ($P < 0.001$). For difficult face mask ventilation, we could only estimate summary sensitivity (0.17, 95% CI 0.06 to 0.39) and specificity (0.90, 95% CI 0.81 to 0.95) for the modified Mallampati test.

Authors' conclusions

Bedside airway examination tests, for assessing the physical status of the airway in adults with no apparent anatomical airway abnormalities, are designed as screening tests. Screening tests are expected to have high sensitivities. We found that all investigated index tests had relatively low sensitivities with high variability. In contrast, specificities were consistently and markedly higher than sensitivities across all tests. The standard bedside airway examination tests should be interpreted with caution, as they do not appear to be good screening tests. Among the tests we examined, the upper lip bite test showed the most favourable diagnostic test accuracy properties. Given the paucity of available data, future research is needed to develop tests with high sensitivities to make them useful, and to consider their use for screening difficult face mask ventilation and failed intubation. The 27 studies in 'Studies awaiting classification' may alter the conclusions of the review, once we have assessed them.

PLAIN LANGUAGE SUMMARY

Bedside examination tests to detect beforehand adults who are likely to be difficult to intubate

Review question

We looked for the most suitable and accurate rapid screening test in adults with no obvious airway abnormalities, to identify those who are likely to be difficult to intubate (i.e. insertion of a tube into the windpipe).

Background

Intubation ensures a patient's airway is clear while they are heavily sedated, unconscious or anaesthetized, so their breathing can be controlled by machine (ventilation), and appropriate levels of oxygen can be given during surgery, following major trauma, during critical illness, or following cardiac arrest. Having an airway that is difficult to intubate is a potentially life-threatening situation.

Tube insertion is preceded by laryngoscopy (insertion of mini-camera to view route of tube insertion), requires advanced skills, and is generally uneventful. Intubation is difficult in approximately 10% of patients, who require special equipment and precautions. Several physical features are associated with difficult airways and failed intubation, so warning of potentially difficult airways would be helpful.

Several quick bedside tests are in routine clinical use to identify those at high risk for difficult airways, but how accurate these are remains unclear.

Population

We included studies of adults aged 16 years or older without obvious airway abnormalities who were to receive standard intubation.

Test under investigation

We assessed the seven most common bedside tests, routinely used to detect difficult airways. These take only a few seconds to complete and require no special equipment.

The index tests (diagnostic tests of interest) included:

- the Mallampati test (original or modified; asking a sitting patient to open his mouth and to protrude the tongue as much as possible so that visibility can be determined);

- Wilson risk score (including patient's weight, head and neck movement, jaw movement, receding chin, buck teeth);

- thyromental distance (length between the chin and the upper edge of Adam's apple);

- sternomental distance (length between the chin and the notch between the collar bones);

- mouth opening test;

- upper lip bite test;

- or any combination of these tests.

Search date

The evidence is current to 16 December 2016. (We searched for new studies in March 2018, but we have not yet included them in the review.)

Study characteristics

We included 133 studies (844,206 participants) which investigated the accuracy of the seven tests above, plus 69 other common tests and 32 test combinations, in detection of difficult airways.

Key results

For difficult laryngoscopy, the average sensitivity (percentage of correctly identified difficult airways) ranged from 22% (mouth opening test) to 63% (upper lip bite test). The average specificity (percentage of correctly classified patients without difficult airways) ranged from 80% (modified Mallampati test) to 95% (Wilson risk score). The upper lip bite test had the highest sensitivity of all tests considered.

For difficult tube insertion, the average sensitivity ranged from 24% (thyromental distance) to 51% (modified Mallampati test) and the average specificity ranged from 87% (modified Mallampati test) to 93% (mouth opening test). The modified Mallampati test had the highest sensitivity of all tests considered.

For difficult face mask ventilation (another indication of a difficult airway), there were only enough data to calculate average sensitivity of 17% and specificity 90% for the modified Mallampati test.

Quality of the evidence

Overall, the evidence from the studies was of moderate to high quality. The likelihood of the studies providing reliable results was generally high, although in half of them, the intubating physician knew the result of the preceding test, which may have influenced results, but this is the normal situation in routine clinical care. The characteristics of patients, tests, and conditions were comparable to those seen in a wide range of everyday clinical settings. The results of this review should apply to standard preoperative airway assessments in apparently normal hospital patients worldwide.

Conclusion

The bedside screening tests examined in this review are not well suited for the purpose of detecting unanticipated difficult airways because they missed a large number of people who had a difficult airway.

SUMMARY OF FINDINGS FOR THE MAIN COMPARISON *[Explanation]*

Patients or population: adults with no apparent anatomical airway abnormalities

Settings: operating theatres, intensive care units and emergency departments

Studies: total of 133 studies, mostly cohort type studies; six case-control studies. Each study can be present in more than one analysis

Test	Number of participants (studies)	Summary sensitivity (95% confidence interval)	Summary specificity (95% confidence interval)	Prevalence median (IQR)	Implications	Quality and comments
Difficult laryngoscopy						
Mallampati test	2165 (6)	0.40 (0.16 to 0.71)	0.89 (0.75 to 0.96)	10% (5% to 16%)	With a prevalence of 10%, 10 out of 100 patients will have difficult laryngoscopy. Of these, 6 will be missed by the Mallampati test (60% of 10). Of the 90 patients without difficult laryngoscopy 10 will be unnecessarily classified as having difficult airway	Limited number of studies. Risk of bias mostly low in all domains. Applicability concerns low.
Modified Mallampati test	232,939 (80)	0.53 (0.47 to 0.59)	0.80 (0.74 to 0.85)	10% (5% to 16%)	With a prevalence of 10%, 10 out of 100 patients will have difficult laryngoscopy. Of these, 5 will be missed by the modified Mallampati test (47% of 10). Of the 90 patients without difficult laryngoscopy 18 will be unnecessar-	Risk of bias mostly unclear in all domains. Applicability concerns mostly low.

					ily classified as having a difficult airway	
Wilson risk score	5862 (5)	0.51 (0.40 to 0.61)	0.95 (0.88 to 0.98)	10% (5% to 16%)	With a prevalence of 10%, 10 out of 100 patients will have difficult laryngoscopy. Of these, 5 will be missed by the Wilson risk score (49% of 10). Of the 90 patients without difficult laryngoscopy 5 will be unnecessarily classified as having a difficult airway	Limited number of studies. Risk of bias unclear. Applicability concerns mostly low in all domains.
Thyromental distance	33,189 (42)	0.37 (0.28 to 0.47)	0.89 (0.84 to 0.93)	10% (5% to 16%)	With a prevalence of 10%, 10 out of 100 patients will have difficult laryngoscopy. Of these, 6 will be missed by thyromental distance (63% of 10). Of the 90 patients without difficult laryngoscopy 10 will be unnecessarily classified as having a difficult airway	Risk of bias mostly low in all domains. Applicability concerns low.
Sternomental distance	12,211 (16)	0.33 (0.16 to 0.56)	0.92 (0.86 to 0.96)	10% (5% to 16%)	With a prevalence of 10%, 10 out of 100 patients will have difficult laryngoscopy. Of these, 7 will be missed by sternomental distance (67% of 10). Of the 90 patients without diffi-	Risk of bias mostly low in all domains. Applicability concerns low.

					cult laryngoscopy 7 will be unnecessarily classified as having a difficult airway	
Mouth opening test	22,179 (24)	0.22 (0.13 to 0.33)	0.94 (0.90 to 0.97)	10% (5% to 16%)	With a prevalence of 10%, 10 out of 100 patients will have difficult laryngoscopy. Of these, 8 will be missed by the mouth opening test (78% of 10). Of the 90 patients without difficult laryngoscopy 5 will be unnecessarily classified as having a difficult airway	Risk of bias mostly low in all domains. Applicability concerns low.
Upper lip bite test	19,609 (27)	0.67 (0.45 to 0.83)	0.92 (0.86 to 0.95)	10% (5% to 16%)	With a prevalence of 10%, 10 out of 100 patients will have difficult laryngoscopy. Of these, 3 will be missed by the upper lip bite test (33% of 10). Of the 90 patients without difficult laryngoscopy 7 will be unnecessarily classified as having a difficult airway	Risk of bias mostly low in all domains. Applicability concerns low.
Difficult tracheal intubation						
Modified Mallampati test	191,849 (24)	0.51 (0.40 to 0.61)	0.87 (0.82 to 0.91)	11% (5% to 13%)	With a prevalence of 11%, 11 out of 100 patients will have difficult tracheal intuba-	Risk of bias mostly unclear in all domains. Applicability concerns mostly low in all do-

					tion. Of these, 5 will be missed by the modified Mallampati test (49% of 11). Of the 89 patients without difficult tracheal intubation 12 will be unnecessarily classified as having a difficult airway	mains.
Thyromental distance	5089 (10)	0.24 (0.12 to 0.43)	0.90 (0.80 to 0.96)	11% (5% to 13%)	With a prevalence of 11%, 11 out of 100 patients will have difficult tracheal intubation. Of these, 8 will be missed by thyromental distance (76% of 11). Of the 89 patients without difficult tracheal intubation 9 will be unnecessarily classified as having a difficult airway	Risk of bias mostly low in all domains. Applicability concerns low.
Mouth opening test	6091 (9)	0.27 (0.16 to 0.41)	0.93 (0.87 to 0.96)	11% (5% to 13%)	With a prevalence of 11%, 11 out of 100 patients will have difficult tracheal intubation. Of these, 8 will be missed by the mouth opening test (73% of 11). Of the 89 patients without difficult tracheal intubation 6 will be unnecessarily classified as having a difficult airway	Risk of bias mostly low in all domains. Applicability concerns low.
Difficult face mask ventilation						

Modified test	Mallampati	56,323 (6)	0.17 (0.06 to 0.39)	0.90 (0.81 to 0.95)	11% (6% to 28%)	With a prevalence of 11%, 11 out of 100 patients will have difficult face mask ventilation. Of these, 9 will be missed by the modified Mallampati test (83% of 11). Of the 89 patients without difficult face mask ventilation 9 will be unnecessarily classified as having a difficult airway	Risk of bias mostly unclear in all domains. Applicability concerns mostly low.

CAUTION: the results on this table should not be interpreted in isolation from the results of the individual included studies contributing to each summary test accuracy measure. We have reported these in the main body of the text of the review. We calculated prevalences from the included studies by reference standard

IQR: interquartile range.

BACKGROUND

Target condition being diagnosed

The difficult airway is a potentially life-threatening event during anaesthesia, following major trauma, with the onset of critical illness, and for resuscitation following cardiac arrest. While any part of the respiratory tract (through which air passes during breathing) is considered to be part of the airway, the difficult airway is focused on the upper airway, that is, the portion of the respiratory tract that extends from the nares or mouth to, and including, the larynx. Thus subglottic stenosis, a type of airway obstruction, and other subglottic problems are not part of this definition of the difficult airway. The upper airway must be patent (open and unblocked) to allow spontaneous lung ventilation by the patient and for physician-, nurse- or therapist-managed assisted ventilation. Additionally, during severe illness or states of altered consciousness the airway must be secured to prevent soiling of the lower airway (trachea, bronchi, etc.) and lung parenchyma by gastric contents, oral secretions, infectious material and blood. Without a patent airway, asphyxia develops within seconds to minutes; without resolution of the loss of a patent airway, death occurs quickly (ASA 2003).

Most commonly, there is an orderly sequence of events in the process of upper airway management by practitioners that terminates with endotracheal intubation. The initial step is the application of a tight fitting face mask with the patient continuing to breath spontaneously. Typically, sedating and paralysing drugs are administered to facilitate airway access. This is followed by the application of positive airway pressure, generated manually with a breathing bag, to provide assisted ventilation. Next, a laryngoscope is inserted into the mouth and pharynx to allow visualization of the glottis and, finally, a tracheal tube is advanced through the glottis into the trachea (ASA 2003).

The difficult airway is not a disease; neither is it just one particular anatomical characteristic of patient physiognomy. Strictly speaking, the difficult airway (or difficult airway event) describes difficulty in or failure to complete one or more of the sequential steps in upper airway management. It is a complex interaction of patient anatomy, clinical circumstances and clinician skill. Nevertheless, the usual focus of the difficult airway is anatomical anomalies in contrast to functional airway obstruction that can accompany inadequate anaesthesia (the struggling patient, coughing, laryngospasm, opioid induced skeletal muscle and laryngeal rigidity, bronchospasm etc.). Thus, the difficult airway does not have a reference standard other than the result of the actual attempted airway management for a patient. While there are no standardized definitions of a difficult airway event, the 2003 practice guidelines from the American Society of Anesthesiologists (ASA), suggested using at least four descriptors of difficult airway events (ASA 2003). In a simplified form, these are as follows.

- Difficult face mask ventilation: it is not possible to provide adequate face mask ventilation.
- Difficult laryngoscopy: even with multiple attempts it is not possible to visualize any portion of the vocal cords during conventional laryngoscopy.
- Difficult tracheal intubation: tracheal intubation requires multiple attempts.
- Failed intubation: placement of the tracheal tube fails after multiple intubation attempts.

Current guidelines added difficult placement or functioning of supraglottic devices as a dimension for the difficult airway (ASA 2013). Difficult face mask ventilation is generally due to an inadequate mask fit or excessive resistance to gas ingress or egress; face mask ventilation is usually facilitated by the insertion of an oral airway or by the administration of muscle relaxants (El-Orbany 2009). Beside the signs of absent or inadequate chest wall movement and breath sounds, difficult ventilation is also recognized by falling oxygen saturation or increasing partial pressure of carbon dioxide in the arterial blood ($PaCO_2$), or both. Kheterpal and colleagues reported the risk of this difficult airway event during anaesthesia in over 50,000 patients (Kheterpal 2009). Using the descriptions "difficult ventilation (inadequate, unstable, or requiring two providers) with or without muscle relaxant" and "unable to mask ventilate with or without muscle relaxant" the prevalence was 2.2% (1141/53,401) for the former and 0.15% (77/53,041) for the latter.

The standard rigid laryngoscope typically consists of a handle containing batteries and an interchangeable blade with a light source. There are many types of laryngoscope blades. The two main types are the curved Macintosh blade and the straight Miller blade. The tip of the Macintosh blade is advanced into the vallecula, where it sits anterior to the epiglottis and raises the epiglottis out of the visual pathway; the Miller blade is advanced further into the airway with the tip sitting posterior to the epiglottis, trapping and elevating the epiglottis while exposing the glottis and vocal folds. However, both a curved and a straight blade can be used in either fashion with the tip of the blade in the vallecula or behind the epiglottis. Each blade comes in several lengths and widths to accommodate patient size (ASA 2003).

Cormack and Lehane (Cormack 1984), proposed a four-grade scoring system to describe the view at direct laryngoscopy. using these standard laryngoscopes. The assigned grades are:

- full view of the glottis;
- partial view of the glottis or arytenoids;
- only epiglottis visible;
- neither glottis nor epiglottis visible.

This scoring system was extended by Yentis and Lee (Yentis 1998), by subdividing grade 2 into 2a) partial view of the glottis; and 2b) arytenoids or posterior part of the vocal cords only just visible. Other modifications of the Cormack and Lehane grades were proposed in the Cook 2000 study. Difficult laryngoscopy is usu-

ally defined as a laryngeal exposure with a score of grade 3 or grade 4. A systematic review (9 studies with 14,438 participants) found that the prevalence of difficult laryngoscopy ranged from 6% to 27% (Lee 2006); in these nine studies the original Cormack and Lehane grades were almost uniformly used to classify difficult laryngoscopy.

Difficult tracheal intubation has been variously defined as a procedure requiring excessive time, multiple attempted passages of the tracheal tube, or having to resort to specialized equipment. A quantitative intubation difficulty scale has been proposed (Adnet 1997). Lundstrom and colleagues defined a non-difficult tracheal intubation as "intubated by direct laryngoscope by the first anaesthetist and in two attempts maximally"; difficult tracheal intubation was any event with more than two anaesthetists, more than two attempts, use of specialized equipment or failed passage (Lundstrom 2009). In a cohort of over 90,000 patients having anaesthesia, the prevalence of difficult tracheal intubation was 5.2% (4704/91,297) (Lundstrom 2009).

Failed intubation is the least common of the difficult airway events. Lundstrom and colleagues reported a frequency of 0.15% in 91,297 participants (Lundstrom 2009). Failed intubation may be defined as "a maximum of three attempts at intubation; a fourth attempt by a more experienced colleague is permissible. If unsuccessful, a failed intubation should be declared and Plan B implemented" (DA Society 2015).

Because the definitions of the difficult airway are not standardized, the prevalence depends on the definition. For example, the Rose 1996 study used three definitions for difficulties during intubation. These are: poor view at laryngoscopy (Cormack and Lehane grade 3 to 4); three or more direct laryngoscopy attempts before insertion of the endotracheal tube; and failure to insert the endotracheal tube with direct laryngoscopy. The prevalence was 10.1%, 1.9%, and 0.1% respectively. The prevalence also depends on the circumstances of medical management, being more difficult in a prehospital setting (Adnet 1997). The Combes 2006 study found the prevalence of difficult tracheal intubation in a prehospital setting to be 7.4%.

Index test(s)

The difficult airway may be the result of obvious upper airway pathology or anatomical anomaly. When such upper airway distortion is obvious, the prudent practitioner will choose alternate plans for airway management. It is the unanticipated difficult airway in a patient without obvious airway pathology or anatomical anomaly that has fostered the search for diagnostic screening tests. These have most commonly been extensions of the physical examination of the patient, with a grading or scoring system for one or more particular attributes of the head, neck and mouth. Some of these particular attributes that are thought to be relevant for detection of the unanticipated difficult airway include the following (ASA 2013).

- Distance between upper and lower incisors.
- Length of the upper incisors.
- Neck length.
- Neck diameter.
- Range of neck flexion and extension.
- Shape of the palate.
- Thyromental distance.
- Tissue compliance of the submandibular space.
- Relationship of maxillary and mandibular incisors during normal jaw closure.
- Relationship of maxillary and mandibular incisors during voluntary protrusion of mandible.
- Visibility of the uvula.

The most popular of these screening tests by airway physical examination include the following.

- Mallampati test.
- Modified Mallampati test.
- Wilson risk score.
- Thyromental distance.
- Sternomental distance.
- Mouth opening test.
- Upper lip bite test.

See Table 1 for more details.

Clinical pathway

Before patients undergo surgery with general anaesthesia, it is common practice to screen for a difficult airway. This screening includes taking a medical history and identifying overt flags for the difficult airway, such as malformations or deformations. For these individuals, alternative methods of airway management are planned in advance. For the remaining apparently normal patients, there is still a risk of unanticipated difficult airway.

To further reduce the number of individuals with an unanticipated difficult airway, clinicians perform bedside airway physical examination tests. The results of these bedside tests help healthcare providers to plan different levels of alternative airway management. A difficult airway occurs in the early phases of general anaesthesia when airway management takes place.

With the exception of the Wilson risk score, each of these tests can be completed in five to 15 seconds; the Wilson risk score also requires information about the patient's weight. Two tests, Mallampati and thyromental distance, have been combined in some reports of screening tests. The performance of these tests by different examiners can have large interobserver variability. The Karkouti 1996 study had two observers independently perform an airway physical examination with 10 characteristics in 59 patients, including some of the specific tests in Table 1. The poorest test performance was with the Mallampati, with classification of patients having only a fair agreement between the observers (Kappa coefficient 0.31). The difficulty in achieving repeatability of airway

classification may explain some of the skepticism about using the index tests before surgery.

Rationale

The serious morbidity and mortality associated with unsuccessfully managed upper airway was recognized decades ago. This prompted the standard use of pulse oximetry and capnography during anaesthesia and emergency care. In addition, learned societies, in particular the ASA and the Difficult Airway Society, have promulgated guidelines for management of the difficult airway (ASA 2013; DA Society 2015). There is indirect evidence from the ASA's closed claims analysis that claims for death and brain damage during the induction of anaesthesia have decreased between the years 1985 to 1992 and 1993 to 1999 (Peterson 2005). Also, the Berkow 2009 study reported a reduction in the need for an emergent surgical airway via tracheostomy through the introduction of a comprehensive difficult airway programme. These improvements in outcomes have been ascribed to standardized airway examination, improved monitoring, new airway devices and technology, and practice guidelines. Specifically, significant advances in the availability of robust video laryngoscopy equipment and other airway devices, such as laryngeal mask airways, have dramatically increased the techniques available for patients with a difficult airway (Luba 2010; Pott 2008).

The role of screening tests and their benefits are still uncertain. Four systematic reviews of airway examination tests have been published (ASA 2003; Lee 2006; Lundstrom 2011; Shiga 2005). The ASA Taskforce concluded that "There is insufficient published evidence to evaluate the predictive value of multiple features of the airway physical examination versus single features in predicting the presence of a difficult airway" and "An airway physical examination should be conducted, whenever feasible, before the initiation of anaesthetic care and airway management in all patients" (ASA 2013); this report did not present a meta-analysis. The Lee 2006 systematic review and meta-analysis reported that "the Mallampati tests have limited accuracy for predicting the difficult airway and thus are not useful screening tests". The Lundstrom 2011 systematic review and meta-analysis was limited to the modified Mallampati score only. Their conclusion was "that the modified Mallampati score is inadequate as a stand-alone test of a difficult laryngoscopy or tracheal intubation". The Shiga 2005 systematic review and meta-analysis of six airway screening tests found that "the clinical value of bedside screening tests for predicting difficult intubation remains limited". Nevertheless, an airway physical examination is still recommended (ASA 2003; ASA 2013). For example, airway examination may be useful in order to select the patients for which newer devices are most likely to be useful.

Since the previous systematic reviews, new statistical methods for the meta-analysis of diagnostic tests with correct handling of the dependency structure of such data are available. For example, the variability of the predictive performance of a diagnostic test in future patients can now be more correctly estimated. Additionally, more studies of large sample size have been published. This review will incorporate an up-to-date literature search and new statistical methods to establish the diagnostic properties of airway physical examination screening tests.

OBJECTIVES

The objective of this review was to characterize and compare the diagnostic accuracy of the Mallampati classification and other commonly used airway examination tests for assessing the physical status of the airway in adult patients with no apparent anatomical airway abnormalities. We performed this individually for each of the four descriptors of the difficult airway: difficult face mask ventilation, difficult laryngoscopy, difficult tracheal intubation, and failed intubation.

METHODS

Criteria for considering studies for this review

Types of studies

We considered diagnostic test accuracy studies (case-control or consecutive series) of any individual index test or a combination of the tests listed in Table 1 against a reference standard. We required studies to provide data for true positives, false positives, false negatives and true negatives. We excluded studies that were reported only in abstract form, were uncontrolled reports (case series, case reports), randomized controlled trials of test-treatment design that are more appropriately analysed as intervention than as diagnostic test accuracy studies, and studies that examined an index test other than bedside tests (for example, those involving radiological imaging).

Participants

We included adults of either sex, aged 16 years or greater, without obvious airway abnormalities who were having laryngoscopy performed with a standard laryngoscope (usually size 3 Macintosh blade) and the trachea intubated with a styletted or non-styletted tracheal tube. We excluded studies performed in populations with a high prevalence of abnormal airways (maxillofacial trauma, cervical spine trauma, or otorhinolaryngology tumours) or those performed using specialized laryngoscopes or techniques (for example, awake fibreoptic intubation).

Index tests

We included bedside tests used singly or in combination for detection of a difficult airway. These include any version of the Mallampati test (Ezri 2001; Mallampati 1985; Samsoon 1987), Wilson risk score (Wilson 1988), thyromental distance (Lewis 1994), sternomental distance (Ramadhani 1996), mouth opening test (Calder 2003), and upper lip bite test (Khan 2003), but were not limited to these tests. We collected information on the inter- or intraobserver correlation of the tests, or both, if reported or referenced in the study.

Target conditions

The target condition was difficult airway. Although the difficult airway does not have a reference standard other than the result of the actual attempted airway management for a patient, the 2003 practice guidelines of the American Society of Anesthesiologists (ASA), suggested using at least four descriptions of difficult airway events (ASA 2003), as follows.

- Difficult face mask ventilation.
- Difficult laryngoscopy.
- Difficult tracheal intubation.
- Failed intubation.

Reference standards

As outline above in Target condition being diagnosed, the reference standards were: difficult face mask ventilation, difficult laryngoscopy, difficult tracheal intubation, and failed intubation. As there were no standard definitions for the reference standards, we accepted the authors' definition used for each study.

Search methods for identification of studies

We performed electronic searches and searched other resources.

Electronic searches

The search is current to 16 December 2016. For identifying any eligible studies, we searched the following electronic databases.

- Cochrane Central Register of Controlled Trials (CENTRAL; 2016, Issue 11), in the Cochrane Library (see Appendix 1).
- Cochrane Register of Diagnostic Test Accuracy Studies
- MEDLINE Ovid SP (1946 to 16 December 2016; see Appendix 2).
- Embase Ovid SP (1874 to 16 December 2016; see Appendix 3).
- ISI Web of Science (1950 to 16 December 2016; see Appendix 4).
- CINAHL EBSCO host (1982 to 16 December 2016; see Appendix 5).

When searching the databases, we used both subject headings and free text terms. We adapted our MEDLINE search strategy for searching all other databases.

We also searched the following regional electronic bibliographic databases, subject-specific databases, and dissertation and theses databases.

- IndMED
- KoreaMED
- LILACS
- Panteleimon
- PASCAL
- Google Scholar
- Turning Research into Practice (TRIP) database
- DissOnline
- OpenSIGLE

We did not apply any language restrictions.

We performed a further search in March 2018. We have added those results to 'Studies awaiting classification' and we will incorporate them into the review at the next update.

Searching other resources

For identifying any additional published, unpublished and ongoing studies, we searched the Science Citation Index and checked the references of all the relevant studies. We also handsearched the following journals and proceedings of the following conferences.

- *Acta Anaesthesiologica Scandinavica* (from 1995 to 29 December 2016).
- *British Journal of Anaesthesia* (from 1995 to 29 December 2016).
- *Canadian Journal of Anesthesia* (from 1995 to 29 December 2016).
- *Critical Care Medicine* (from 1995 to 29 December 2016).
- *Intensive Care Medicine* (from 1995 to 29 December 2016).
- *American Journal of Respiratory and Critical Care Medicine* (from 1995 to 29 December 2016).
- Abstracts from congresses of the European Society of Anaesthesiology (from 2004 to 29 December 2016).
- Abstracts from the International Anesthesia Research Society (from 2000 to 29 December 2016).
- ATS international conference proceedings (from 2008 to 29 December 2016).
- International Symposium on Intensive Care and Emergency Medicine proceedings (from 1997 to 29 December 2016).
- American Society of Anesthesiologists Annual Meeting proceedings (from 2000 to 29 December 2016).

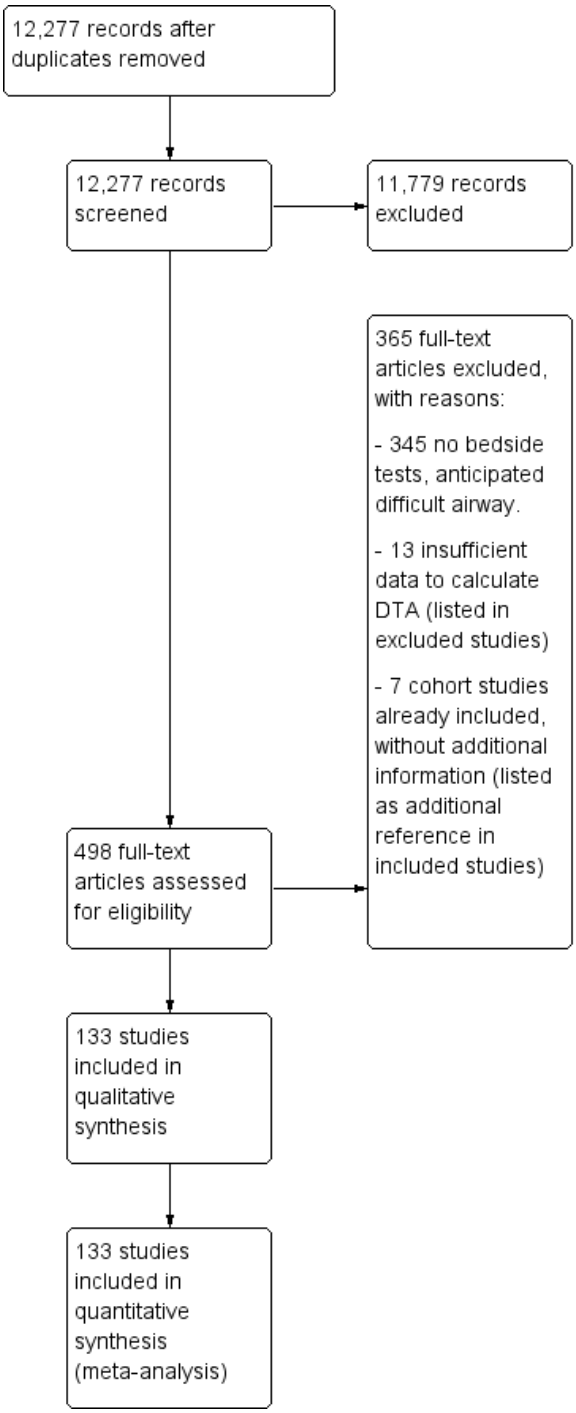
We also searched guidelines by the French, Italian, Spanish and German Societies of Anaesthesiology and Intensive Care.

Data collection and analysis

Selection of studies

NLP, DR and HH independently, and in duplicate, performed selection of studies. We resolved disagreements by discussion or by involving AL as arbiter. We initially screened studies by the title and abstract and then retrieved full reports for potentially relevant studies. For these studies, we used a predefined electronic spreadsheet to assess and document studies for inclusion and exclusion according to the above selection criteria. We documented study selection in a flow chart (Liberati 2009; Figure 1).

Figure 1. Study flow diagram.



Data extraction and management

We independently, and in duplicate, performed data extraction using a predefined electronic spreadsheet within the database, MS Access. We resolved disagreements by discussion or by involving AL or HH as arbiter. We then transferred data to Review Manager 5 (Review Manager 2014), Stata 14 (Stata 2015) and to R (R 2017), for further calculations.

Assessment of methodological quality

We independently, and in duplicate, performed assessment of methodological quality using a predefined electronic spreadsheet. We resolved disagreements by discussion or by involving AL or HH as arbiter. We used all four domains (Table 2), from the QUADAS-2 tool (Whiting 2011), a revision of the original QUADAS tool (Whiting 2003), to assess the methodological quality of the included studies that is implemented in Review Manager 2014. This included the risk of bias with signalling questions and applicability judgement. We presented both a description and the judgement (coded 'yes', 'no', or 'unclear') for each signalling question. Additionally, we coded risk of bias and applicability as 'high', 'low', or 'unclear'.

We piloted the quality checklist independently on a sample of five papers and refined the checklist before proceeding further. When necessary, we contacted authors of original studies for information on unclear quality items.

We have presented the items on methodological quality assessments in methodological quality summary figuress 12 to 15 in the Results section. In addition, we have presented methodological quality graphs showing the relative distribution of methodological quality assessments for each included study in Appendix 6.

Statistical analysis and data synthesis

For each included study, we treated the index test results as separate binary classifiers; we recorded the cutpoint for dichotomization. The included studies reported one or more difficult airway events. We separately tallied each type of reported difficult airway event. We collected details on definitions of positive and negative reference standard responses. We constructed 2×2 tables of test and reference standard results to show the cross-classification of difficult airway status and test outcome. In studies where multiple index tests were performed, we also constructed a series of 2×2 tables where the results of investigations were combined, provided that they were derived from the total study population, and that the definition of a positive result for combined tests was reported.

We used sensitivity and specificity of each test or test combination as the underlying parameter in our calculations. As healthcare providers want to avoid false negatives, we considered sensitivity as the most important property when comparing diagnostic accuracy between tests: overlooking a person at high risk for a difficult airway event may be potentially life-threatening during anaesthesia. False positives on the other hand, have less severe implications in this scenario. To describe and visualize the data, we produced forest plots showing pairs of sensitivity and specificity together with 95% confidence intervals (CIs) from each study in Review Manager 2014. We presented data for all eligible studies on forest plots, but included only cohort type studies in the meta-analyses to minimise the risk of bias. We meta-analysed pairs of sensitivity and specificity using a generalized linear mixed model approach to perform a bivariate meta-analysis of sensitivity and specificity (Chu 2006).

We primarily performed meta-analyses for pooling estimates using the 'lme4' package in R (R 2017). From this package we used the bivariate binomial method using the glmer function. We presented results as sensitivity and specificity, as from the bivariate estimates (logit transformed) with 95% CIs.

We produced a specificity versus sensitivity plot showing the study estimates of individual studies, the summary receiver operating characteristic (ROC) point (summary values for sensitivity and specificity) and the 95% confidence region around the summary ROC point.

We indirectly compared index tests and index test combinations by including a covariate for test type in bivariate models (i.e. meta-regression) using methods suggested by Partlett and Takwoingi (Partlett 2016). For pairwise, between-index-test difference comparisons, we used a bivariate mixed effects regression model to test the joint null hypothesis of no difference in sensitivity and specificity between two index tests as calculated in the models described above. We formally compared models using a likelihood ratio test. If we rejected the joint null hypothesis, we individually compared sensitivity and specificity. We present differences only for test comparison pairs where sufficient data were available and where models converged.

Investigations of heterogeneity

To explore heterogeneity, we considered patient demographics (e.g. age, sex, weight); the indication for airway management (e.g. elective surgery, emergent surgery, critical illness, trauma, resuscitation); and different standards for declaring a difficult airway as potential covariates in a bivariate model (Whiting 2011).

Sensitivity analyses

We assessed the impact of study design on our findings by excluding case-control studies. We assessed the impact of the risk of bias

due to lack of blinding by excluding studies where the results of the index tests were not blinded.

Assessment of reporting bias

Testing for reporting bias and small study effects may not be especially useful in the context of studies of diagnostic tests (Begg 2005), therefore, we did not present analyses on reporting bias.

RESULTS

Results of the search

We searched up to 16 December 2016. Our search yielded a total of 12,277 papers after combining search results from all sources and after removing duplications. Based on independent title and abstract evaluations, we excluded 11,779 references and retrieved the full text for 498 references. After careful evaluation, we excluded another 365 studies (Figure 1). The studies we excluded because of insufficient data are reported in the 'Characteristics of excluded studies' tables. After detailed assessments, we included 133 studies involving 844,206 participants (Figure 1).

From an updated search in March 2018, we have added 27 study reports to 'Characteristics of studies awaiting classification' tables.

The 133 studies evaluated a total of seven different prespecified test strategies, as well as 69 non-prespecified, and 32 combinations (Table 3). For the prespecified index tests, we found six studies for the Mallampati test, 105 for the modified Mallampati test, six for the Wilson risk score, 52 for thyromental distance, 18 for sternomental distance, 34 for the mouth opening test and 30 for the upper lip bite test. A total of 42 studies evaluated one individual test, 36 studies evaluated two tests, 21 studies evaluated three tests, and 36 studies evaluated four to 12 tests. Eberhart 2005 reported interobserver correlation (IOC) for the upper lip bite test (IOC = 0.79), and for the modified Mallampati test (IOC = 0.59). None of the studies reported intraobserver correlations.

Table 1 defines the cut-off thresholds for index tests. Details on reported cut-offs are presented in the 'Characteristics of included studies' tables. Overall, we did not consider variations of cut-offs to be clinically important.

Eighteen comparisons (7 studies) defined the target condition as: difficult face mask ventilation; 218 comparisons (92 studies) as difficult laryngoscopy; 72 comparisons (50 studies) as difficult tracheal intubation; and two comparisons (two studies) as failed intubation.

The median number of participants per study was 380, with an interquartile range (IQR) from 200 to 662. The median (IQR) percentage of females included in the studies was 53% (44% to 64%). The median (IQR) age of the participants was 45 years (39 to 52). The median (IQR) body mass index (BMI) was 27.3 kg/m^2(24.8 to 30.0). All studies, apart from two (Freund 2012; Soyuncu 2009), performed airway management in the operating theatre; the Freund 2012 study involved airway management in ambulance cars and the Soyuncu 2009 study in an emergency department. We did not subgrouped studies according to where the study took place. The characteristics of the individual studies are described in the 'Characteristics of included studies' tables.

Methodological quality of included studies

We report the details for individual study quality in the 'Characteristics of included studies' tables. Due to the complex structure of the review (multiple combinations of index tests and reference standards reported within individual studies), we were not able to use Review Manager 5 in-built features to report all the risk of bias domains and applicability concerns for each study in the 'Characteristics of included studies' tables (Review Manager 2014).

We judged the risk of bias to be variable, across all studies, for the different domains; with mostly low risk of bias observed with patient selection, flow and timing, and mostly unclear risk of bias with reference standard and index test. We judged applicability concerns to be low for all domains. Most of the included studies were cohort type studies, only six of the included studies were case-control type studies (Connor 2011; Frerk 1996; Fritscherova 2011; Naguib 1999; Naguib 2006; Nath 1997). Given the nature of the setting, and the test, we did not observe partial or differential verification in any of the studies.

For difficult laryngoscopy, the reference standard was performed blinded in 42, non-blinded in six, and blinding was unclear in 43 studies. For difficult tracheal intubation, the reference standard was performed blinded in 11, non-blinded in eight, and blinding was unclear in 29 studies. For difficult face mask ventilation, the reference standard was performed blinded in one, non-blinded in one and blinding was unclear in five studies. For failed intubation, the reference standard was performed blinded in none, non-blinded in none and blinding was unclear in three studies. The index test was blinded in all studies investigating prespecified index tests as expected. Among alternative tests or test combinations, eight studies had non-blinded index tests (Fritscherova 2011; Gonzalez 2008; Hagiwara 2015; Kim 2011; Langeron 2000; Nath 1997; Wilson 1988; Wong 1999). All studies evaluated the index test before the reference standard, except for the Fritscherova 2011 study, which performed the index test the day after intubation. Ninety-five studies included all participants in the analysis. We found incomplete or unclear reporting in 40 studies.

For a graphical summary of the risk of bias and applicability, see the graphs in Appendix 6. For a summary for each difficult airway component, refer to Figure 2 for difficult laryngoscopy; Figure 3 for difficult tracheal intubation; Figure 4 for failed intubation; and Figure 5 for difficult face mask ventilation.

Figure 2. Risk of bias and applicability concerns graph for difficult laryngoscopy: review authors' judgements about each domain presented as percentages across included studies.

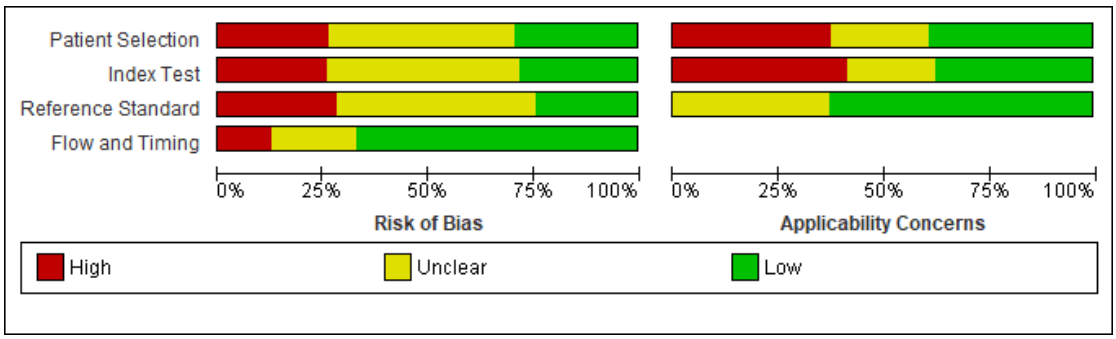


Figure 3. Risk of bias and applicability concerns graph for difficult tracheal intubation: review authors' judgements about each domain presented as percentages across included studies.

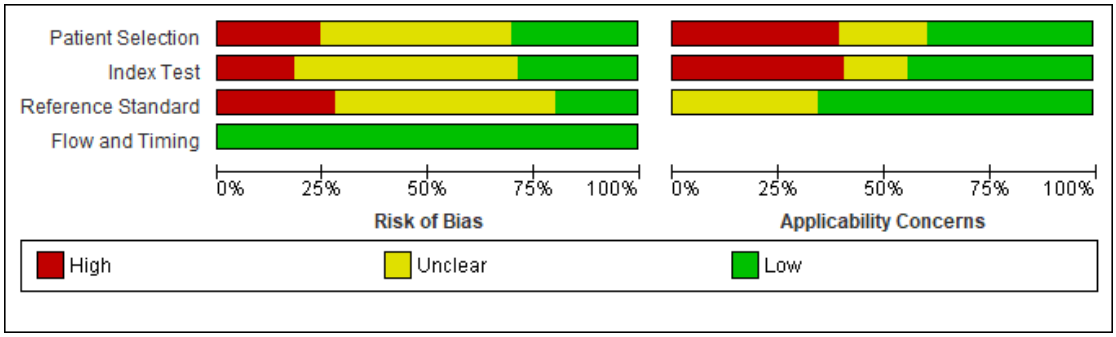


Figure 4. Risk of bias and applicability concerns graph for failed intubation: review authors' judgements about each domain presented as percentages across included studies.

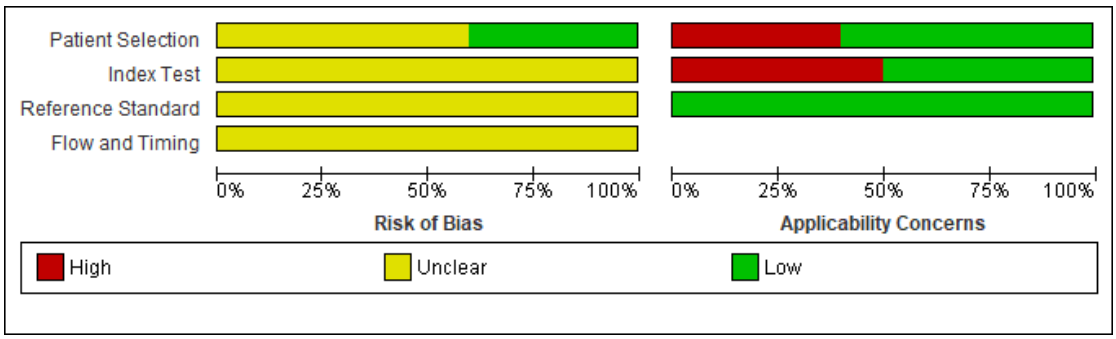
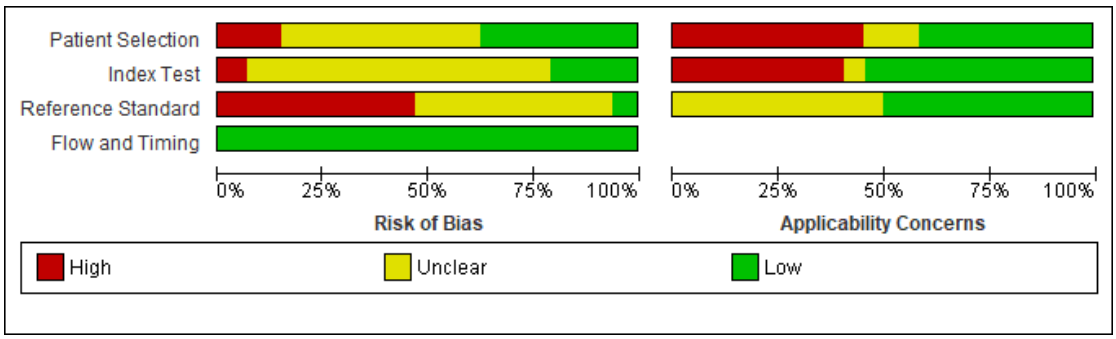


Figure 5. Risk of bias and applicability concerns graph for difficult face mask ventilation: review authors' judgements about each domain presented as percentages across included studies.



Findings

The median (IQR) prevalence for difficult laryngoscopy, difficult tracheal intubation, difficult face mask ventilation, and failed intubation was 11% (6% to 19%), 13% (5% to 16%), 6% (5% to 25%) and 0.6% (0.3% to 0.9%), respectively.

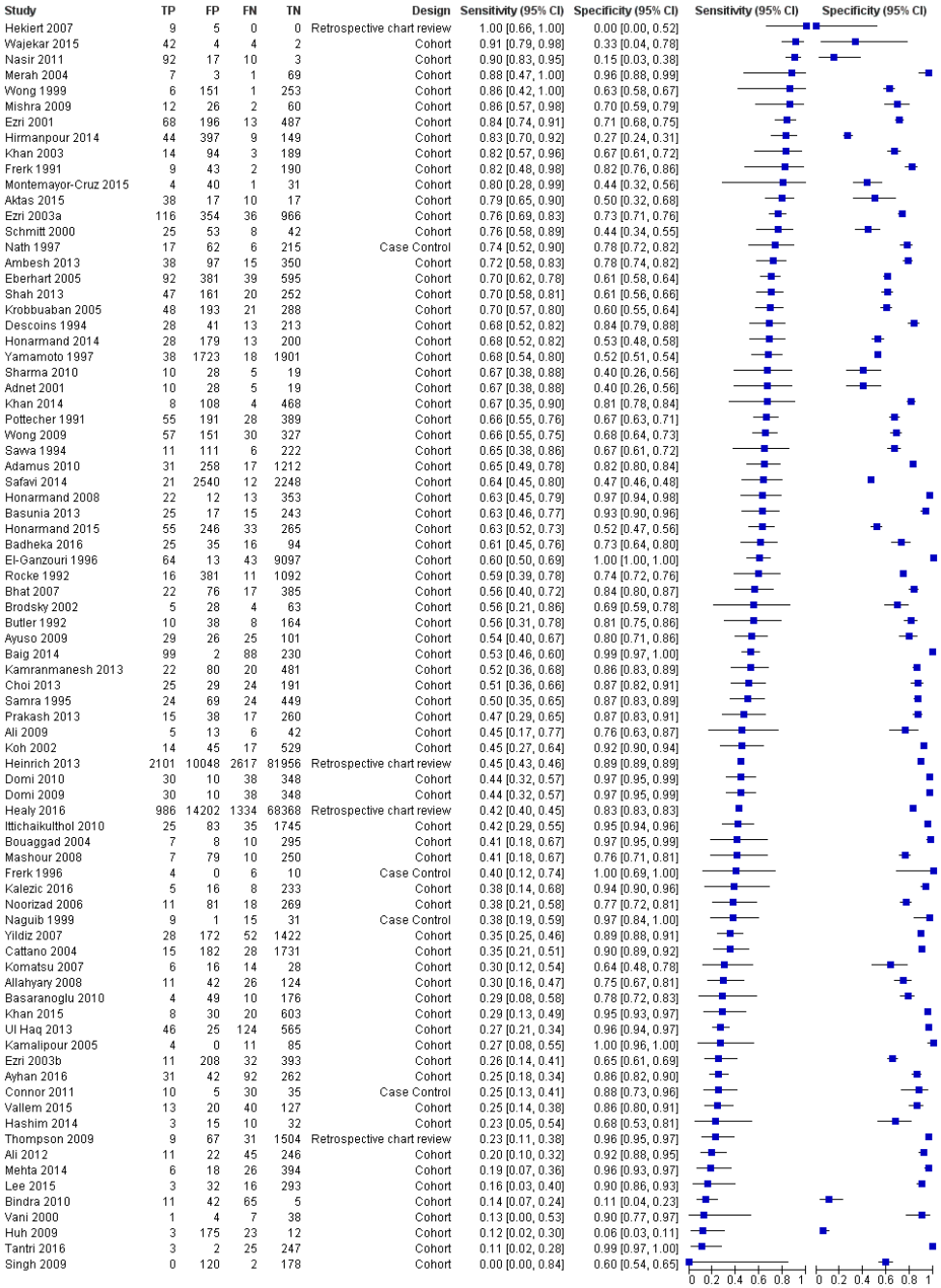
We were able to perform meta-analyses for 11 comparisons (all 7 prespecified index tests for difficult laryngoscopy; modified Mallampati test, thyromental distance and mouth opening test for difficult tracheal intubation; modified Mallampati test for difficult face mask ventilation). We did not perform meta-analyses of studies with the Mallampati test, Wilson risk score, sternomental distance or upper lip bite test for difficult tracheal intubation; studies with thyromental distance, upper lip bite test or mouth opening test for difficult face mask ventilation; or studies with the modified Mallampati test for failed intubation because only one or two studies were available. For the remaining comparisons, we did not find any studies. All studies that we included in the meta-analyses used one clinically identical cut-off value per test. See Summary of findings for key findings.

Difficult laryngoscopy

For the Mallampati test, there were six studies involving 2165 participants with 153 cases of difficult laryngoscopy (Data table 1). Sensitivity varied from 0.05 to 0.85, and specificity from 0.65 to 0.98. We estimated a summary sensitivity of 0.40 (95% confidence interval (CI) 0.16 to 0.71) and a summary specificity of 0.89 (95% CI 0.75 to 0.96).

For the modified Mallampati test, there were 80 studies involving 232,939 participants with 10,545 cases of difficult laryngoscopy (Data table 3). Both sensitivity and specificity varied from 0.00 to 1.00. We estimated a summary sensitivity of 0.53 (95% CI 0.47 to 0.59) and a summary specificity of 0.80 (95% CI 0.74 to 0.85). See Figure 6.

Figure 6. Forest plot of modified Mallampati test for difficult laryngoscopy, sorted by descending sensitivity. Summary sensitivity 0.53 (95% confidence interval (CI) 0.47 to 0.59); summary specificity 0.80 (95% CI 0.74 to 0.85).

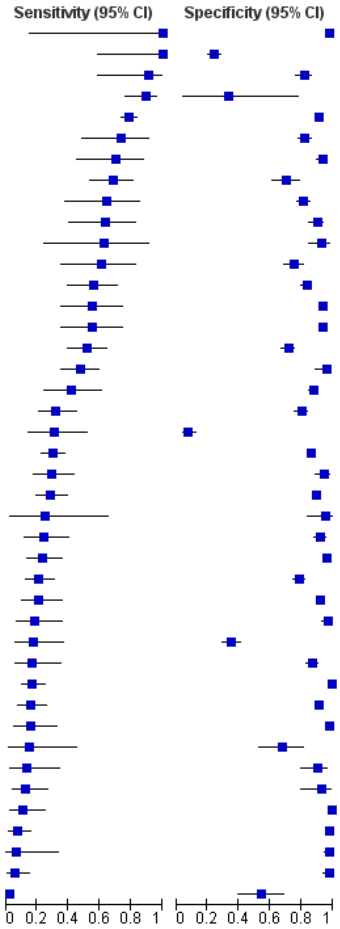


Study	TP	FP	FN	TN	Design	Sensitivity (95% CI)	Specificity (95% CI)
Hekiert 2007	9	5	0	0	Retrospective chart review	1.00 [0.66, 1.00]	0.00 [0.00, 0.52]
Wajekar 2015	42	4	4	2	Cohort	0.91 [0.79, 0.98]	0.33 [0.04, 0.78]
Nasir 2011	92	17	10	3	Cohort	0.90 [0.83, 0.95]	0.15 [0.03, 0.38]
Merah 2004	7	3	1	69	Cohort	0.88 [0.47, 1.00]	0.96 [0.88, 0.99]
Wong 1999	6	151	1	253	Cohort	0.86 [0.42, 1.00]	0.63 [0.58, 0.67]
Mishra 2009	12	26	2	60	Cohort	0.86 [0.57, 0.98]	0.70 [0.59, 0.79]
Ezri 2001	68	196	13	487	Cohort	0.84 [0.74, 0.91]	0.71 [0.68, 0.75]
Hirmanpour 2014	44	397	9	149	Cohort	0.83 [0.70, 0.92]	0.27 [0.24, 0.31]
Khan 2003	14	94	3	189	Cohort	0.82 [0.57, 0.96]	0.67 [0.61, 0.72]
Frerk 1991	9	43	2	190	Cohort	0.82 [0.48, 0.98]	0.82 [0.76, 0.86]
Montemayor-Cruz 2015	4	40	1	31	Cohort	0.80 [0.28, 0.99]	0.44 [0.32, 0.56]
Aktas 2015	38	17	10	17	Cohort	0.79 [0.65, 0.90]	0.50 [0.32, 0.68]
Ezri 2003a	116	354	36	966	Cohort	0.76 [0.69, 0.83]	0.73 [0.71, 0.76]
Schmitt 2000	25	53	8	42	Cohort	0.76 [0.58, 0.89]	0.44 [0.34, 0.55]
Nath 1997	17	62	6	215	Case Control	0.74 [0.52, 0.90]	0.78 [0.72, 0.82]
Ambesh 2013	38	97	15	350	Cohort	0.72 [0.58, 0.83]	0.78 [0.74, 0.82]
Eberhart 2005	92	381	39	595	Cohort	0.70 [0.62, 0.78]	0.61 [0.58, 0.64]
Shah 2013	47	161	20	252	Cohort	0.70 [0.58, 0.81]	0.61 [0.56, 0.66]
Krobbuaban 2005	48	193	21	288	Cohort	0.70 [0.57, 0.80]	0.60 [0.55, 0.64]
Descoins 1994	28	41	13	213	Cohort	0.68 [0.52, 0.82]	0.84 [0.79, 0.88]
Honarmand 2014	28	179	13	200	Cohort	0.68 [0.52, 0.82]	0.53 [0.48, 0.58]
Yamamoto 1997	38	1723	18	1901	Cohort	0.68 [0.54, 0.80]	0.52 [0.51, 0.54]
Sharma 2010	10	28	5	19	Cohort	0.67 [0.38, 0.88]	0.40 [0.26, 0.56]
Adnet 2001	10	28	5	19	Cohort	0.67 [0.38, 0.88]	0.40 [0.26, 0.56]
Khan 2014	8	108	4	468	Cohort	0.67 [0.35, 0.90]	0.81 [0.78, 0.84]
Pottecher 1991	55	191	28	389	Cohort	0.66 [0.55, 0.76]	0.67 [0.63, 0.71]
Wong 2009	57	151	30	327	Cohort	0.66 [0.55, 0.75]	0.68 [0.64, 0.73]
Sawa 1994	11	111	6	222	Cohort	0.65 [0.38, 0.86]	0.67 [0.61, 0.72]
Adamus 2010	31	258	17	1212	Cohort	0.65 [0.49, 0.78]	0.82 [0.80, 0.84]
Safavi 2014	21	2540	12	2248	Cohort	0.64 [0.45, 0.80]	0.47 [0.46, 0.48]
Honarmand 2008	22	12	13	353	Cohort	0.63 [0.45, 0.79]	0.97 [0.94, 0.98]
Basunia 2013	25	17	15	243	Cohort	0.63 [0.46, 0.77]	0.93 [0.90, 0.96]
Honarmand 2015	55	246	33	265	Cohort	0.63 [0.52, 0.73]	0.52 [0.47, 0.56]
Badheka 2016	25	35	16	94	Cohort	0.61 [0.45, 0.76]	0.73 [0.64, 0.80]
El-Ganzouri 1996	64	13	43	9097	Cohort	0.60 [0.50, 0.69]	1.00 [1.00, 1.00]
Rocke 1992	16	381	11	1092	Cohort	0.59 [0.39, 0.78]	0.74 [0.72, 0.76]
Bhat 2007	22	76	17	385	Cohort	0.56 [0.40, 0.72]	0.84 [0.80, 0.87]
Brodsky 2002	5	28	4	63	Cohort	0.56 [0.21, 0.86]	0.69 [0.59, 0.78]
Butler 1992	10	38	8	164	Cohort	0.56 [0.31, 0.78]	0.81 [0.75, 0.86]
Ayuso 2009	29	26	25	101	Cohort	0.54 [0.40, 0.67]	0.80 [0.71, 0.86]
Baig 2014	99	2	88	230	Cohort	0.53 [0.46, 0.60]	0.99 [0.97, 1.00]
Kamranmanesh 2013	22	80	20	481	Cohort	0.52 [0.36, 0.68]	0.86 [0.83, 0.89]
Choi 2013	25	29	24	191	Cohort	0.51 [0.36, 0.66]	0.87 [0.82, 0.91]
Samra 1995	24	69	24	449	Cohort	0.50 [0.35, 0.65]	0.87 [0.83, 0.89]
Prakash 2013	15	38	17	260	Cohort	0.47 [0.29, 0.65]	0.87 [0.83, 0.91]
Ali 2009	5	13	6	42	Cohort	0.45 [0.17, 0.77]	0.76 [0.63, 0.87]
Koh 2002	14	45	17	529	Cohort	0.45 [0.27, 0.64]	0.92 [0.90, 0.94]
Heinrich 2013	2101	10048	2617	81956	Retrospective chart review	0.45 [0.43, 0.46]	0.89 [0.89, 0.89]
Domi 2010	30	10	38	348	Cohort	0.44 [0.32, 0.57]	0.97 [0.95, 0.99]
Domi 2009	30	10	38	348	Cohort	0.44 [0.32, 0.57]	0.97 [0.95, 0.99]
Healy 2016	986	14202	1334	68368	Retrospective chart review	0.42 [0.40, 0.45]	0.83 [0.83, 0.83]
Ittichaikulthol 2010	25	83	35	1745	Cohort	0.42 [0.29, 0.55]	0.95 [0.94, 0.96]
Bouaggad 2004	7	8	10	295	Cohort	0.41 [0.18, 0.67]	0.97 [0.95, 0.99]
Mashour 2008	7	79	10	250	Cohort	0.41 [0.18, 0.67]	0.76 [0.71, 0.81]
Frerk 1996	4	0	6	10	Case Control	0.40 [0.12, 0.74]	1.00 [0.69, 1.00]
Kalezic 2016	5	16	8	233	Cohort	0.38 [0.14, 0.68]	0.94 [0.90, 0.96]
Noorizad 2006	11	81	18	269	Cohort	0.38 [0.21, 0.58]	0.77 [0.72, 0.81]
Naguib 1999	9	1	15	31	Case Control	0.38 [0.19, 0.59]	0.97 [0.84, 1.00]
Yildiz 2007	28	172	52	1422	Cohort	0.35 [0.25, 0.46]	0.89 [0.88, 0.91]
Cattano 2004	15	182	28	1731	Cohort	0.35 [0.21, 0.51]	0.90 [0.89, 0.92]
Komatsu 2007	6	16	14	28	Cohort	0.30 [0.12, 0.54]	0.64 [0.48, 0.78]
Allahyary 2008	11	42	26	124	Cohort	0.30 [0.16, 0.47]	0.75 [0.67, 0.81]
Basaranoglu 2010	4	49	10	176	Cohort	0.29 [0.08, 0.58]	0.78 [0.72, 0.83]
Khan 2015	8	30	20	603	Cohort	0.29 [0.13, 0.49]	0.95 [0.93, 0.97]
Ul Haq 2013	46	25	124	565	Cohort	0.27 [0.21, 0.34]	0.96 [0.94, 0.97]
Kamalipour 2005	4	0	11	85	Cohort	0.27 [0.08, 0.55]	1.00 [0.96, 1.00]
Ezri 2003b	11	208	32	393	Cohort	0.26 [0.14, 0.41]	0.65 [0.61, 0.69]
Ayhan 2016	31	42	92	262	Cohort	0.25 [0.18, 0.34]	0.86 [0.82, 0.90]
Connor 2011	10	5	30	35	Case Control	0.25 [0.13, 0.41]	0.88 [0.73, 0.96]
Vallem 2015	13	20	40	127	Cohort	0.25 [0.14, 0.38]	0.86 [0.80, 0.91]
Hashim 2014	3	15	10	32	Cohort	0.23 [0.05, 0.54]	0.68 [0.53, 0.81]
Thompson 2009	9	67	31	1504	Retrospective chart review	0.23 [0.11, 0.38]	0.96 [0.95, 0.97]
Ali 2012	11	22	45	246	Cohort	0.20 [0.10, 0.32]	0.92 [0.88, 0.95]
Mehta 2014	6	18	26	394	Cohort	0.19 [0.07, 0.36]	0.96 [0.93, 0.97]
Lee 2015	3	32	16	293	Cohort	0.16 [0.03, 0.40]	0.90 [0.86, 0.93]
Bindra 2010	11	42	65	5	Cohort	0.14 [0.07, 0.24]	0.11 [0.04, 0.23]
Vani 2000	1	4	7	38	Cohort	0.13 [0.00, 0.53]	0.90 [0.77, 0.97]
Huh 2009	3	175	23	12	Cohort	0.12 [0.02, 0.30]	0.06 [0.03, 0.11]
Tantri 2016	3	2	25	247	Cohort	0.11 [0.02, 0.28]	0.99 [0.97, 1.00]
Singh 2009	0	120	2	178	Cohort	0.00 [0.00, 0.84]	0.60 [0.54, 0.65]

For the Wilson risk score, there were five studies involving 5862 participants with 145 cases of difficult laryngoscopy (Data table 7). Sensitivity varied from 0.00 to 0.75, and specificity from 0.86 to 0.99. We estimated a summary sensitivity of 0.51 (95% CI 0.40 to 0.61) and a summary specificity of 0.95 (95% CI 0.88 to 0.98). For thyromental distance, there were 42 studies involving 33,189 participants with 2364 cases of difficult laryngoscopy (Data table 9). Sensitivity varied from 0.03 to 1.00, and specificity from 0.08 to 0.99. We estimated a summary sensitivity of 0.37 (95% CI 0.28 to 0.47) and a summary specificity of 0.89 (95% CI 0.84 to 0.93). See Figure 7.

Figure 7. Forest plot of thyromental distance for difficult laryngoscopy, sorted by descending sensitivity. Summary sensitivity 0.37 (95% CI 0.28 to 0.47); summary specificity 0.89 (95% CI 0.84 to 0.93).

Study	TP	FP	FN	TN	Design	Sensitivity (95% CI)	Specificity (95% CI)
Singh 2009	2	6	0	292	Cohort	1.00 [0.16, 1.00]	0.98 [0.96, 0.99]
Wong 1999	7	306	0	98	Cohort	1.00 [0.59, 1.00]	0.24 [0.20, 0.29]
Frerk 1991	10	43	1	190	Cohort	0.91 [0.59, 1.00]	0.82 [0.76, 0.86]
Wajekar 2015	41	4	5	2	Cohort	0.89 [0.76, 0.96]	0.33 [0.04, 0.78]
Khan 2013	209	387	56	3848	Cohort	0.79 [0.73, 0.84]	0.91 [0.90, 0.92]
Khan 2009a	14	64	5	297	Cohort	0.74 [0.49, 0.91]	0.82 [0.78, 0.86]
Salimi 2008	14	22	6	305	Cohort	0.70 [0.46, 0.88]	0.93 [0.90, 0.96]
Badheka 2016	35	35	16	84	Cohort	0.69 [0.54, 0.81]	0.71 [0.62, 0.79]
Sawa 1994	11	62	6	271	Cohort	0.65 [0.38, 0.86]	0.81 [0.77, 0.85]
Konwar 2015	14	18	8	160	Cohort	0.64 [0.41, 0.83]	0.90 [0.84, 0.94]
Merah 2004	5	5	3	67	Cohort	0.63 [0.24, 0.91]	0.93 [0.85, 0.98]
Butler 1992	11	50	7	152	Cohort	0.61 [0.36, 0.83]	0.75 [0.69, 0.81]
Ezri 2003b	24	101	19	500	Cohort	0.56 [0.40, 0.71]	0.83 [0.80, 0.86]
Domi 2010	15	25	12	374	Cohort	0.56 [0.35, 0.75]	0.94 [0.91, 0.96]
Domi 2009	15	25	12	374	Cohort	0.56 [0.35, 0.75]	0.94 [0.91, 0.96]
Krobbuaban 2005	36	136	33	345	Cohort	0.52 [0.40, 0.64]	0.72 [0.67, 0.76]
Fritscherova 2011	35	3	39	71	Case Control	0.47 [0.36, 0.59]	0.96 [0.89, 0.99]
Koh 2002	13	71	18	503	Cohort	0.42 [0.25, 0.61]	0.88 [0.85, 0.90]
Tse 1995	20	82	42	327	Cohort	0.32 [0.21, 0.45]	0.80 [0.76, 0.84]
Huh 2009	8	172	18	15	Cohort	0.31 [0.14, 0.52]	0.08 [0.05, 0.13]
Ezri 2003a	46	184	106	1136	Cohort	0.30 [0.23, 0.38]	0.86 [0.84, 0.88]
Ayuso 2009	16	7	38	120	Cohort	0.30 [0.18, 0.44]	0.94 [0.89, 0.98]
Yildiz 2007	23	175	57	1419	Cohort	0.29 [0.19, 0.40]	0.89 [0.87, 0.91]
Vani 2000	2	2	6	40	Cohort	0.25 [0.03, 0.65]	0.95 [0.84, 0.99]
Descoins 1994	10	20	31	234	Cohort	0.24 [0.12, 0.40]	0.92 [0.88, 0.95]
Ittichaikulthol 2010	14	71	46	1757	Cohort	0.23 [0.13, 0.36]	0.96 [0.95, 0.97]
Wong 2009	18	103	67	377	Cohort	0.21 [0.13, 0.31]	0.79 [0.75, 0.82]
Cattano 2004	9	153	34	1760	Cohort	0.21 [0.10, 0.36]	0.92 [0.91, 0.93]
Qudaisat 2011	6	7	26	196	Cohort	0.19 [0.07, 0.36]	0.97 [0.93, 0.99]
Tantri 2016	5	161	23	88	Cohort	0.18 [0.06, 0.37]	0.35 [0.29, 0.42]
Noorizad 2006	5	46	24	304	Cohort	0.17 [0.06, 0.36]	0.87 [0.83, 0.90]
El-Ganzouri 1996	18	99	89	10301	Cohort	0.17 [0.10, 0.25]	0.99 [0.99, 0.99]
Freund 2012	11	52	59	531	Cohort	0.16 [0.08, 0.26]	0.91 [0.88, 0.93]
Mehta 2014	5	11	27	433	Cohort	0.16 [0.05, 0.33]	0.98 [0.96, 0.99]
Hashim 2014	2	15	11	32	Cohort	0.15 [0.02, 0.45]	0.68 [0.53, 0.81]
Nadal 1998	3	6	19	55	Cohort	0.14 [0.03, 0.35]	0.90 [0.80, 0.96]
Connor 2011	5	3	35	37	Case Control	0.13 [0.04, 0.27]	0.93 [0.80, 0.98]
Allahyary 2008	4	1	33	165	Cohort	0.11 [0.03, 0.25]	0.99 [0.97, 1.00]
Shah 2013	5	8	62	404	Cohort	0.07 [0.02, 0.17]	0.98 [0.96, 0.99]
Basaranoglu 2010	1	5	13	220	Cohort	0.07 [0.00, 0.34]	0.98 [0.95, 0.99]
Vallem 2015	3	3	50	144	Cohort	0.06 [0.01, 0.16]	0.98 [0.94, 1.00]
Bilgin 1998	14	22	438	26	Cohort	0.03 [0.02, 0.05]	0.54 [0.39, 0.69]

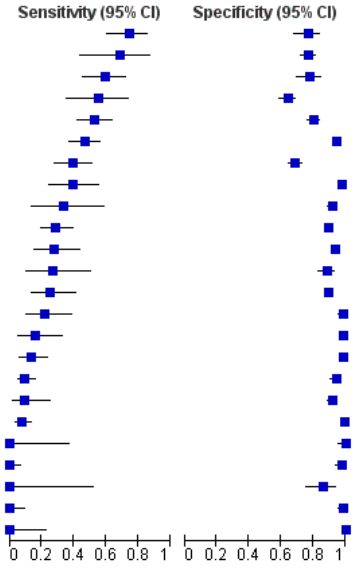


For sternomental distance, there were 16 studies involving 12,211 participants with 762 cases of difficult laryngoscopy (Data table 12). Sensitivity varied from 0.00 to 0.84, and specificity from 0.71 to 1.00. We estimated a sensitivity of 0.33 (95% CI 0.16 to 0.56) and a specificity of 0.92 (95% CI 0.86 to 0.96).

For the mouth opening test, there were 24 studies involving 22,179 participants with 1220 cases of difficult laryngoscopy (Data table 14). Sensitivity varied from 0.00 to 0.75, and specificity from 0.64 to 1.00. We estimated a summary sensitivity of 0.22 (95% CI 0.13 to 0.33) and a summary specificity of 0.94 (95% CI 0.90 to 0.97). See Figure 8.

Figure 8. Forest plot of mouth opening for difficult laryngoscopy, sorted by descending sensitivity. Summary sensitivity 0.22 (95% CI 0.13 to 0.33); summary specificity 0.94 (95% CI 0.90 to 0.97).

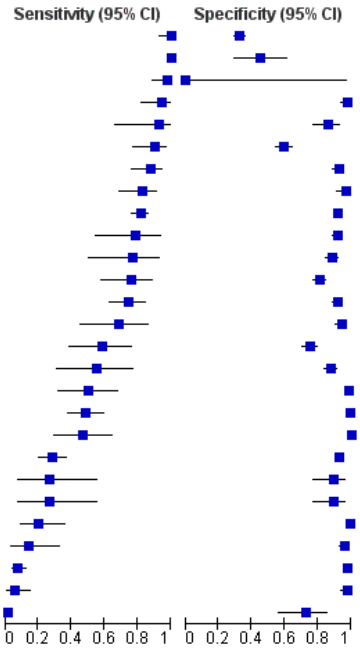
Study	TP	FP	FN	TN	Design	Sensitivity (95% CI)	Specificity (95% CI)
Badheka 2016	38	28	13	91	Cohort	0.75 [0.60, 0.86]	0.76 [0.68, 0.84]
Khan 2009a	13	83	6	278	Cohort	0.68 [0.43, 0.87]	0.77 [0.72, 0.81]
Ayuso 2009	32	28	22	99	Cohort	0.59 [0.45, 0.72]	0.78 [0.70, 0.85]
Nasiri 2013	16	137	13	244	Cohort	0.55 [0.36, 0.74]	0.64 [0.59, 0.69]
Wong 2009	45	96	40	384	Cohort	0.53 [0.42, 0.64]	0.80 [0.76, 0.83]
El-Ganzouri 1996	50	625	57	9775	Cohort	0.47 [0.37, 0.57]	0.94 [0.94, 0.94]
Krobbuaban 2005	27	150	42	331	Cohort	0.39 [0.28, 0.52]	0.69 [0.64, 0.73]
Descoins 1994	16	5	25	249	Cohort	0.39 [0.24, 0.55]	0.98 [0.95, 0.99]
Domi 2009	6	34	12	374	Cohort	0.33 [0.13, 0.59]	0.92 [0.89, 0.94]
Yildiz 2007	23	175	57	1419	Cohort	0.29 [0.19, 0.40]	0.89 [0.87, 0.91]
Cattano 2004	12	130	31	1783	Cohort	0.28 [0.15, 0.44]	0.93 [0.92, 0.94]
Konwar 2015	6	21	16	157	Cohort	0.27 [0.11, 0.50]	0.88 [0.83, 0.93]
Ezri 2003b	11	63	32	538	Cohort	0.26 [0.14, 0.41]	0.90 [0.87, 0.92]
Breckwoldt 2011	8	4	28	236	Cohort	0.22 [0.10, 0.39]	0.98 [0.96, 1.00]
Prakash 2013	5	5	27	293	Cohort	0.16 [0.05, 0.33]	0.98 [0.96, 0.99]
Shah 2013	9	7	58	407	Cohort	0.13 [0.06, 0.24]	0.98 [0.97, 0.99]
Ayhan 2016	12	18	111	286	Cohort	0.10 [0.05, 0.16]	0.94 [0.91, 0.96]
Mehta 2014	3	35	29	382	Cohort	0.09 [0.02, 0.25]	0.92 [0.89, 0.94]
Ezri 2003a	12	8	140	1312	Cohort	0.08 [0.04, 0.13]	0.99 [0.99, 1.00]
Merah 2004	0	0	8	72	Cohort	0.00 [0.00, 0.37]	1.00 [0.95, 1.00]
Vallem 2015	0	3	53	144	Cohort	0.00 [0.00, 0.07]	0.98 [0.94, 1.00]
Montemayor-Cruz 2015	0	8	5	51	Cohort	0.00 [0.00, 0.52]	0.86 [0.75, 0.94]
Allahyary 2008	0	2	37	164	Cohort	0.00 [0.00, 0.09]	0.99 [0.96, 1.00]
Basaranoglu 2010	0	0	14	225	Cohort	0.00 [0.00, 0.23]	1.00 [0.98, 1.00]



For the upper lip bite test, there were 27 studies involving 19,609 participants with 1998 cases of difficult laryngoscopy (Data table 17). Sensitivity varied from 0.02 to 1.00, and specificity from 0.00 to 1.00. We estimated a summary sensitivity of 0.67 (95% CI 0.45 to 0.83) and a summary specificity of 0.92 (95% CI 0.86 to 0.95). See Figure 9.

Figure 9. Forest plot of upper lip bite test for difficult laryngoscopy, sorted by descending sensitivity. Summary sensitivity 0.67 (95% CI 0.45 to 0.83); summary specificity 0.92 (95% CI 0.86 to 0.95).

Study	TP	FP	FN	TN	Design	Sensitivity (95% CI)	Specificity (95% CI)
Hirmanpour 2014	53	601	0	296	Cohort	1.00 [0.93, 1.00]	0.33 [0.30, 0.36]
Kolarkar 2015	260	22	0	18	Cohort	1.00 [0.99, 1.00]	0.45 [0.29, 0.62]
Wajekar 2015	45	1	1	0	Cohort	0.98 [0.88, 1.00]	0.00 [0.00, 0.97]
Allahyary 2008	35	4	2	162	Cohort	0.95 [0.82, 0.99]	0.98 [0.94, 0.99]
Mishra 2009	13	12	1	74	Cohort	0.93 [0.66, 1.00]	0.86 [0.77, 0.93]
Honarmand 2014	37	155	4	226	Cohort	0.90 [0.77, 0.97]	0.59 [0.54, 0.64]
Ali 2012	49	19	7	249	Cohort	0.88 [0.76, 0.95]	0.93 [0.89, 0.96]
Badheka 2016	42	4	9	115	Cohort	0.82 [0.69, 0.92]	0.97 [0.92, 0.99]
Khan 2013	216	360	49	3875	Cohort	0.82 [0.76, 0.86]	0.91 [0.91, 0.92]
Khan 2009a	15	30	4	331	Cohort	0.79 [0.54, 0.94]	0.92 [0.88, 0.94]
Khan 2003	13	32	4	251	Cohort	0.76 [0.50, 0.93]	0.89 [0.84, 0.92]
Safavi 2014	25	85	8	359	Cohort	0.76 [0.58, 0.89]	0.81 [0.77, 0.84]
Shah 2013	50	35	17	378	Cohort	0.75 [0.63, 0.84]	0.92 [0.88, 0.94]
Konwar 2015	15	10	7	178	Cohort	0.68 [0.45, 0.86]	0.95 [0.90, 0.97]
Nasiri 2013	17	97	12	291	Cohort	0.59 [0.39, 0.76]	0.75 [0.70, 0.79]
Salimi 2008	11	40	9	290	Cohort	0.55 [0.32, 0.77]	0.88 [0.84, 0.91]
Mehta 2014	16	6	16	411	Cohort	0.50 [0.32, 0.68]	0.99 [0.97, 0.99]
Honarmand 2015	43	3	45	504	Cohort	0.49 [0.38, 0.60]	0.99 [0.98, 1.00]
Khan 2011	16	0	18	266	Cohort	0.47 [0.30, 0.65]	1.00 [0.99, 1.00]
Eberhart 2005	37	73	94	903	Cohort	0.28 [0.21, 0.37]	0.93 [0.91, 0.94]
Adnet 2001	4	5	11	42	Cohort	0.27 [0.08, 0.55]	0.89 [0.77, 0.96]
Sharma 2010	4	5	11	42	Cohort	0.27 [0.08, 0.55]	0.89 [0.77, 0.96]
Bhat 2007	8	4	31	457	Cohort	0.21 [0.09, 0.36]	0.99 [0.98, 1.00]
Seo 2012	4	12	24	265	Cohort	0.14 [0.04, 0.33]	0.96 [0.93, 0.98]
Myneni 2010	14	157	159	5669	Cohort	0.08 [0.04, 0.13]	0.97 [0.97, 0.98]
Vallem 2015	3	3	50	144	Cohort	0.06 [0.01, 0.16]	0.98 [0.94, 1.00]
Honarmand 2008	6	11	354	29	Cohort	0.02 [0.01, 0.04]	0.72 [0.56, 0.85]

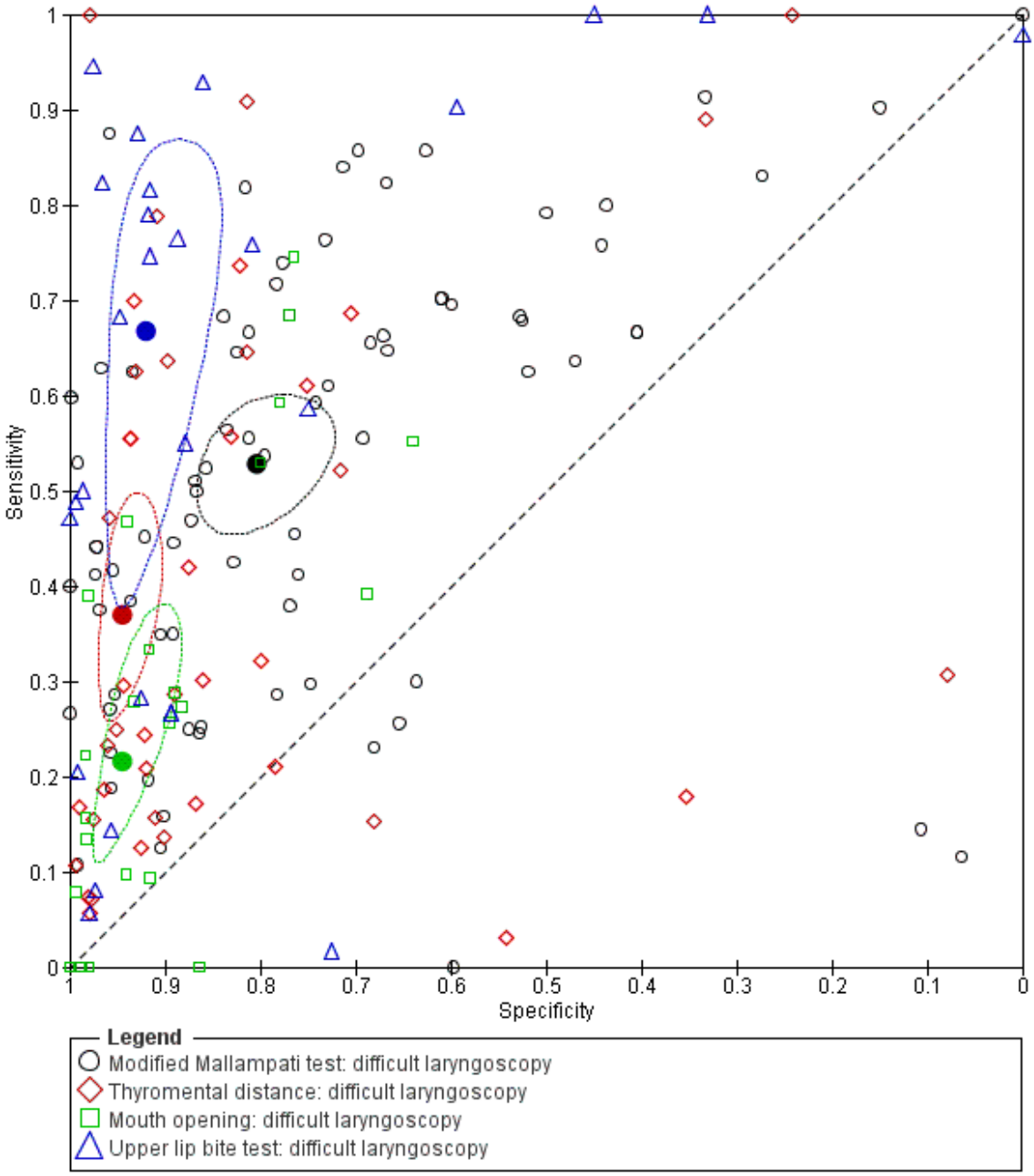


Forty-two studies reported non-prespecified index tests or index test combinations involving 230,680 participants with 7197 cases of difficult laryngoscopy (Data table 20). Both sensitivity and specificity varied from 0.00 to 1.00. We did not perform a meta-analysis on these combinations, as outlined above.

We were able to formally compare four index tests for difficult laryngoscopy. The upper lip bite test had the highest sensitivity, which was significantly different from mouth opening ($P < 0.001$). The modified Mallampati test showed a significantly higher sensitivity compared to thyromental distance ($P = 0.012$) and mouth opening ($P < 0.001$).

Mouth opening had the highest specificity, which was significantly different from the modified Mallampati test ($P < 0.001$). The modified Mallampati test had significantly lower specificity than the upper lip bite test ($P = 0.007$), and thyromental distance ($P = 0.037$). See Figure 10 for a graphical display. We were unable to calculate test comparisons for other test combinations given the lack of data.

Figure 10. Summary receiver operating characteristic (ROC) plot of modified Mallampati test, thyromental distance, mouth opening, and upper lip bite test for difficult laryngoscopy. For each index test, the summary point with the 95% confidence region is displayed.



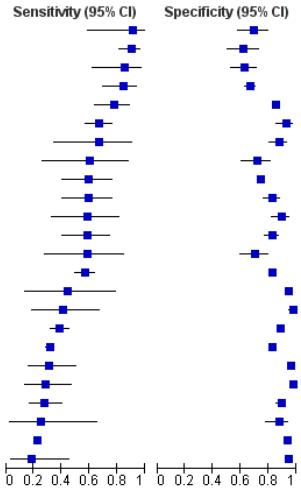
Difficult tracheal intubation

For the Mallampati test, there was only one study (500 participants) with 40 cases of difficult tracheal intubation (Data table 2). Sensitivity in this study was 0.42 (95% CI 0.27 to 0.59), and specificity was 0.93 (95% CI 0.90 to 0.95).

For the modified Mallampati test, there were 24 studies involving 191,849 participants with 6615 cases of difficult tracheal intubation (Data table 5). Sensitivity varied from 0.19 to 0.91, and specificity from 0.62 to 0.98. We estimated a summary sensitivity of 0.51 (95% CI 0.40 to 0.61) and a summary specificity of 0.87 (95% CI 0.82 to 0.91). See Figure 11.

Figure 11. Forest plot of modified Mallampati test for difficult tracheal intubation, sorted by descending sensitivity. Summary sensitivity 0.51 (95% CI 0.40 to 0.61); summary specificity 0.87 (95% CI 0.82 to 0.91).

Study	TP	FP	FN	TN	Design	Sensitivity (95% CI)	Specificity (95% CI)
Knudsen 2014	10	23	1	53	Cohort	0.91 [0.59, 1.00]	0.70 [0.58, 0.80]
Fritscherova 2011	67	28	7	46	Case Control	0.91 [0.81, 0.96]	0.62 [0.50, 0.73]
Juvin 2003	17	41	3	68	Cohort	0.85 [0.62, 0.97]	0.62 [0.53, 0.71]
Pottecher 1991	33	209	6	415	Cohort	0.85 [0.69, 0.94]	0.67 [0.63, 0.70]
Arne 1998	39	168	11	982	Cohort	0.78 [0.64, 0.88]	0.85 [0.83, 0.87]
Naguib 2006	65	7	32	90	Case Control	0.67 [0.57, 0.76]	0.93 [0.86, 0.97]
Gonzalez 2008	8	15	4	104	Cohort	0.67 [0.35, 0.90]	0.87 [0.80, 0.93]
Bergler 1997	6	23	4	58	Cohort	0.60 [0.26, 0.88]	0.72 [0.60, 0.81]
Rocke 1992	19	378	13	1090	Cohort	0.59 [0.41, 0.76]	0.74 [0.72, 0.76]
Tuzuner-Oncul 2008	19	30	13	144	Cohort	0.59 [0.41, 0.76]	0.83 [0.76, 0.88]
Kim 2011	10	11	7	95	Cohort	0.59 [0.33, 0.82]	0.90 [0.82, 0.95]
Seo 2012	21	47	15	222	Cohort	0.58 [0.41, 0.74]	0.83 [0.77, 0.87]
Brodsky 2002	7	26	5	62	Cohort	0.58 [0.28, 0.85]	0.70 [0.60, 0.80]
De Jong 2015	111	346	85	1673	Cohort	0.57 [0.49, 0.64]	0.83 [0.81, 0.84]
Khan 2015	4	34	5	618	Cohort	0.44 [0.14, 0.79]	0.95 [0.93, 0.96]
Ezri 2003c	7	8	10	295	Cohort	0.41 [0.18, 0.67]	0.97 [0.95, 0.99]
Uribe 2015	87	435	138	3482	Cohort	0.39 [0.32, 0.45]	0.89 [0.88, 0.90]
Healy 2016	287	14901	624	69764	Retrospective chart review	0.32 [0.28, 0.35]	0.82 [0.82, 0.83]
K Nasa 2014	10	22	22	530	Cohort	0.31 [0.16, 0.50]	0.96 [0.94, 0.97]
Bouaggad 2004	9	6	23	282	Cohort	0.28 [0.14, 0.47]	0.98 [0.96, 0.99]
Chaves 2009	18	49	47	398	Retrospective chart review	0.28 [0.17, 0.40]	0.89 [0.86, 0.92]
Mallat 2010	2	9	6	63	Cohort	0.25 [0.03, 0.65]	0.88 [0.78, 0.94]
Lundstrom 2009	1042	5898	3620	81531	Cohort	0.22 [0.21, 0.24]	0.93 [0.93, 0.93]
Dohrn 2015	3	24	13	431	Cohort	0.19 [0.04, 0.46]	0.95 [0.92, 0.97]

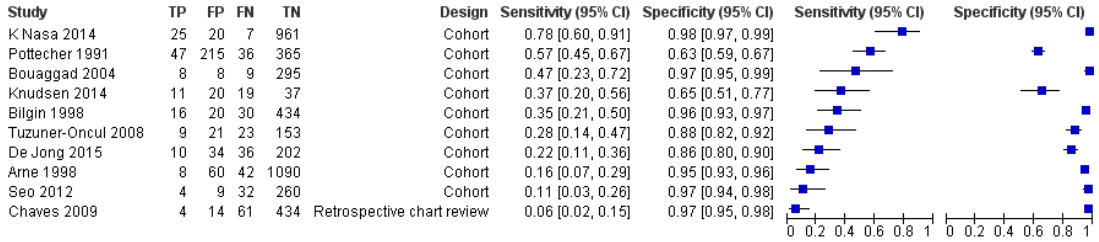


For the Wilson risk score, there was only one study (123 participants) with 17 cases of difficult tracheal intubation (Data table 8). Sensitivity in this study was 0.47 (95% CI 0.23 to 0.72), and specificity was 0.92 (95% CI 0.84 to 0.96).

For sternomental distance, there were two studies (864 participants) with 115 cases of difficult tracheal intubation (Data table 13). Sensitivity varied from 0.31 to 0.60, and specificity from 0.63 to 0.90. We did not perform a meta-analysis on these combinations, as outlined above.

For thyromental distance, there were 10 studies involving 5089 participants with 437 cases of difficult tracheal intubation (Data table 11). Sensitivity varied from 0.06 to 0.78, and specificity from 0.63 to 0.98. We estimated a summary sensitivity of 0.24 (95% CI 0.12 to 0.43) and a summary specificity of 0.90 (95% CI 0.80 to 0.96). See Figure 12.

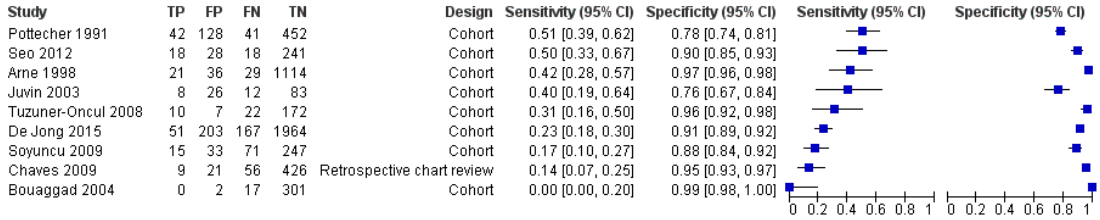
Figure 12. Forest plot of thyromental distance for difficult tracheal intubation, sorted by descending sensitivity. Summary sensitivity 0.24 (95% CI0.12 to 0.43); summary specificity 0.90 (95% CI 0.80 to 0.96).



For the upper lip bite test, there were two studies (598 participants) with 121 cases of difficult tracheal intubation (Data table 19). Sensitivity varied from 0.34 to 0.91, and specificity from 0.93 to 0.96. We did not perform a meta-analysis on these combinations, as outlined above.

For mouth opening, there were 9 studies involving 6091 participants with 607 cases of difficult tracheal intubation (Data table 16). Sensitivity varied from 0.00 to 0.51, and specificity from 0.76 to 0.99. We estimated a summary sensitivity of 0.27 (95% CI 0.16 to 0.41) and a summary specificity of 0.93 (95% CI 0.87 to 0.96). See Figure 13.

Figure 13. Forest plot of mouth opening for difficult tracheal intubation, sorted by descending sensitivity. Summary sensitivity 0.27 (95% CI 0.16 to 0.41); summary specificity 0.93 (95% CI 0.87 to 0.96).

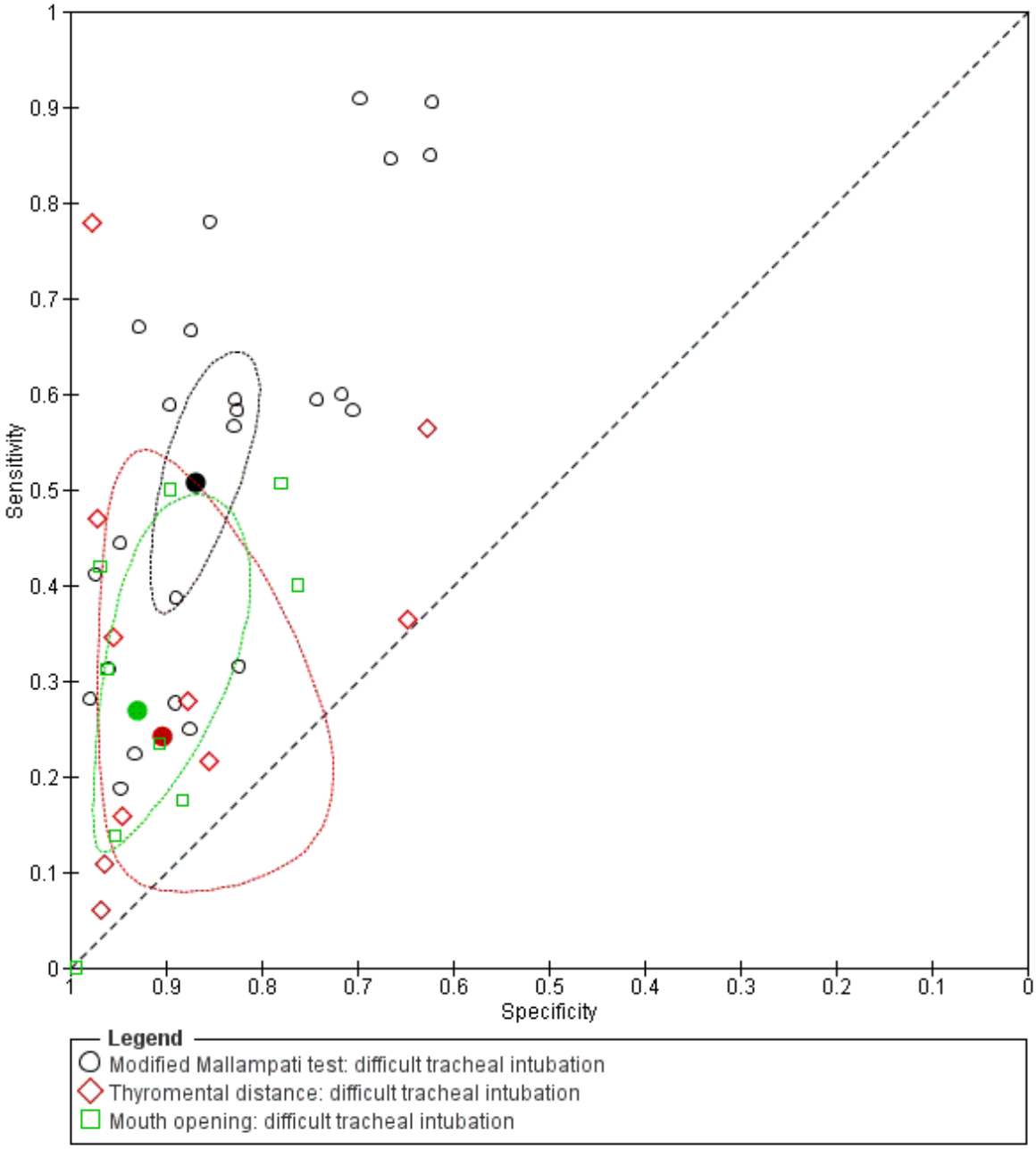


Fifteen studies reported non-prespecified index tests or index test combinations involving 11,089 participants with 1030 cases of difficult tracheal intubation (Data table 22). Sensitivity varied from 0.00 to 0.92, and specificity from 0.48 to 1.00. We did not perform a meta-analysis on these combinations, as outlined above.

We were able to formally compare three index tests for difficult tracheal intubation. The modified Mallampati test had the highest sensitivity. It was significantly higher than the mouth opening test ($P < 0.001$) and thyromental distance ($P < 0.001$). Sensitivity was not significantly different between mouth opening and thyromental distance ($P = 0.07$).

The mouth opening test showed the highest specificity, which was higher than the thyromental distance and the modified Mallampati test. Specificity was significantly different for all test comparisons ($P < 0.001$). See Figure 14 for a graphical display. We were unable to calculate test comparisons for other test combinations, given the lack of data.

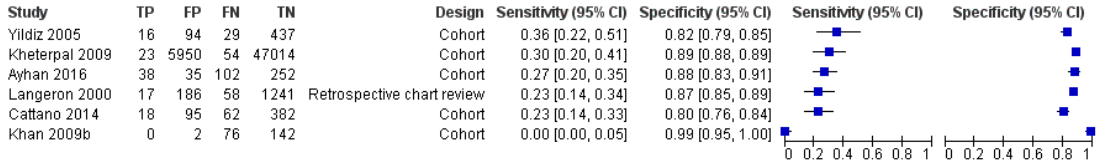
Figure 14. Summary receiver operating characteristic (ROC) plot of modified Mallampati test, thyromental distance, and mouth opening test for difficult tracheal intubation. For each index test the summary point with the 95% confidence region is displayed.



Difficult face mask ventilation

For the modified Mallampati test, there were six studies involving 56,323 participants with 493 cases of difficult face mask ventilation (Data table 4). Sensitivity varied from 0.00 to 0.36, and specificity from 0.80 to 0.99. We estimated a summary sensitivity of 0.17 (95% CI 0.06 to 0.39) and a summary specificity of 0.90 (95% CI 0.81 to 0.95). See Figure 15.

Figure 15. Forest plot of modified Mallampati test for difficult face mask ventilation, sorted by descending sensitivity. Summary sensitivity 0.17 (95% CI 0.06 to 0.39); summary specificity 0.90 (95% CI 0.81 to 0.95).



For thyromental distance, there was only one study (53,041 participants) with 77 cases of difficult face mask ventilation (Data table 10). Sensitivity in this study was 0.13 (95% CI 0.06 to 0.23), and specificity was 0.94 (95% CI 0.94 to 0.95).

For the upper lip bite test, there was only one study (200 participants) with 56 cases of difficult face mask ventilation (Data table 18). Sensitivity in this study was 0.75 (95% CI 0.62 to 0.86), and specificity was 0.60 (95% CI 0.51 to 0.68).

For mouth opening, there were two studies (53,469 participants) with 370 cases of difficult face mask ventilation (Data table 15). Sensitivity was 0.06 in both studies, and specificity ranged from 0.91 to 0.96. We did not perform a meta-analysis on these combinations, as outlined above.

Four studies reported non-prespecified index tests or index test combinations (10,819 participants) with 655 cases of difficult face mask ventilation (Data table 21). Sensitivity varied from 0.04 to 0.81, and specificity from 0.27 to 0.97. We did not perform a meta-analysis on these combinations, as outlined above.

Failed intubation

For the modified Mallampati test, there were two studies (485 participants) with three cases of failed intubation (Data table 6). Sensitivity was 0.00 in one study and not estimable due to a lack of cases (reference standard positives) in the other study. Specificity varied from 0.80 to 0.94. We did not perform a meta-analysis on these combinations, as outlined above.

Sensitivity analyses

We performed sensitivity analyses for study design and lack of blinding of index test results; we found no effect on our findings.

Heterogeneity

For non-prespecified index tests or index test combinations we did not perform a meta-analysis because of the large clinical heterogeneity in terms of differences in test properties. For all index tests where pooling was possible, we found high variability in the estimates.

DISCUSSION

Summary of main results

There was limited to moderate accuracy in commonly used airway examination tests for assessing the physical status of the airway in adult patients with no apparent anatomical airway abnormality. There was a consistent pattern of wide variability in the ranges around the 50% sensitivity point. On the other hand, specificity was high with less variability across most of the tests. This applied likewise for all reference standards. Standard airway examination tests do not appear to work well as screening tests. The potential

high rate of false negatives could lead to disastrous situations during induction of anaesthesia.

Overall, the quality of the estimates was moderate to high. The methodological quality was high for applicability and moderate to high for the risk of bias in the individual studies.

Among the tests under investigation (Summary of findings), the upper lip bite test had the highest sensitivity to foresee difficult laryngoscopy and was significantly better than the modified Mallampati test and the mouth opening test. For difficult tracheal intubation, there was insufficient information for the upper lip bite test. Here the modified Mallampati test had the highest sensitivity. For difficult tracheal intubation, there was no evidence of a difference in sensitivity between the mouth opening and thyromental distance tests. For face mask ventilation and failed intubation, there was insufficient information for test comparisons.

Strengths and weaknesses of the review

This review systematically summarized current evidence about standard bedside airway examination tests using up-to-date methodology from a total of 133 studies involving 844,206 patients. It updates the evidence described in a published systematic review (Lee 2006), and expands the scope of index tests beyond the Mallampati test. We attempted to conduct a comprehensive search for studies, but the fact that 27 studies have not yet been incorporated may be a source of potential bias. We designed our review to cover the most common bedside tests used in clinical routine practice globally. However, this resulted in a large number of comparisons, with varying numbers of studies with sufficient data. We therefore, can provide good quality evidence for a selected set of tests. Moreover, we found a large number of studies on test combinations which contained considerable heterogeneity, and prevented pooling in some cases. We also found some heterogeneity in the definition of index tests and target conditions, which might potentially result in a loss in precision in the estimates. In addition, we were not able to formally analyse the heterogeneity by demographics and the clinical setting such as anaesthesia, critically ill patient, major trauma, or cardiac arrest, where clinicians face very different conditions, sometimes with serious limitations to perform bedside tests.

The risk of bias in the studies, one aspect of quality of the evidence, was generally low. However, as expected, we noted an issue with blinding of the index test results when assessing the target condition in approximately half of the included studies, as in the clinical setting. Despite standardized outcome assessment instruments, this could explain the relatively high specificity compared to sensitivity if outcome assessors tended to classify the airway more frequently difficult when they knew that the bedside test predicted a difficult airway. However, this potential bias may also act in the opposite direction, i.e. better preparation due to knowledge of a potential difficult airway, leading to less problems in actual airway management. We therefore do not expect that this provides sufficient explanation for our results.

Applicability of findings to the review question

The included studies were generally performed in a broad range of standard clinical settings and are expected to apply to standard preoperative airway assessments done in apparently normal hospital patients internationally. This review covers a broad range of standard and routinely applied bedside tests. The outcomes comply with routine target conditions, such as difficult laryngoscopy and difficult tracheal intubation, which all healthcare professionals in the field are familiar with (ASA 2003). For some relevant target conditions, such as difficult face mask ventilation and failed intubation, data were too scarce to draw robust conclusions, and therefore the applicability is limited. As prespecified, this review includes only studies with clinical reference standards, such as difficult tracheal intubation or difficult face mask ventilation. We did not include studies deriving prediction tools solely from radiological imaging and other non-clinical reference standards.

AUTHORS' CONCLUSIONS

Implications for practice

Bedside airway examination tests for assessing the physical status of the airway in adult patients with no apparent anatomical airway abnormality are designed as screening tests. Screening tests are expected to have high sensitivities and depend less on specificity. We found that all the investigated examination tests had relatively low sensitivities with high variability. In contrast, specificities were consistently and markedly better than sensitivities across all tests. Standard airway examination tests do not appear to work well as screening tests. Although false negatives can result from bedside examination tests, it is important to put the risk of an unanticipated difficult airway into context. Whereas failure to predict both difficult face mask ventilation and difficult tracheal intubation could lead to disastrous clinical situations ("cannot intubate - cannot ventilate"), unexpected isolated difficult laryngoscopy might be handled by face mask ventilation. Among the investigated tests, the upper lip bite test showed the most favourable diagnostic test accuracy properties.

The 27 studies in 'Studies awaiting classification' may alter the conclusions of the review once we have assessed them.

Implications for research

Current bedside tests have limited accuracy. Therefore, research to develop tests with high sensitivities are needed to make them useful screening tests. Scarce information is available for difficult

face mask ventilation and failed intubation, which are suitable target conditions to examine in future studies.

ACKNOWLEDGEMENTS

We would like to thank Jane Cracknell, Cochrane Anaesthesia Critical and Emergency Care Group (ACE) Managing Editor; Andrew Smith, ACE Content Editor; Peer Reviewers (Pierre Diemunsch, Haldun Akoglu, Subrahmanyan Radhakrishna); Janet Wale, Consumer Editor; Janne Vendt, Information Specialist who reran the search in December 2016 and March 2018; the Diagnostic Test Accuracy (DTA) Editorial team and Peer Reviewers; and Marialena Trivella, ACE Statistical Editor.

We would like to thank Alexander Hartmann who helped with data extraction for several papers. We thank Pablo Emilio Verde, Senior Researcher, University of Duesseldorf, Germany, for consultation about Bayesian methods. We would also like to thank Yemisi Takwoingi, School of Health and Population Sciences, University of Birmingham, for her help with the GLMM models in R. We would like to thank Zahid Hussain Khan, Professor of Anaesthesiology and Intensive Care at Tehran University of Medical Sciences, Iran, for providing additional information on study patients.

We would also like to thank Mathew Zacharias, ACE Content Editor; Mariska Leeflang, DTA Contact Editor; Pierre Diemunsch and Markus Weiss, ACE Peer Reviewers; and the DTA Peer Reviewers for their help and editorial advice during the preparation of the protocol for the review (Lee 2010).

This review was selected for the third Cochrane Review Support Programme.

REFERENCES

References to studies included in this review

Adamus 2010 *{published data only}*
Adamus M, Fritscherova S, Hrabalek L, Gabrhelik T, Zapletalova J, Janout V. Mallampati test as a predictor of laryngoscopic view. *Biomedical papers of the Medical Faculty of the University Palacký, Olomouc, Czechoslovakia* 2010;**154**(4):339–43. PUBMED: 21293546]

Adnet 2001 *{published data only}*
Adnet F, Racine SX, Borron SW, Clemessy JL, Fournier JL, Lapostolle F, et al. A survey of tracheal intubation difficulty in the operating room: a prospective observational study. *Acta Anaesthesiologica Scandinavica* 2001;**45**(3):327–32. PUBMED: 11207469]

Aktas 2015 *{published data only}*
Aktas S, Atalay YO, Tugrul M. Predictive value of bedside tests for difficult intubations. *European Review for Medical and Pharmacological Sciences* 2015;**19**(9):1595–9. PUBMED: 26004598]

Ali 2009 *{published data only}*
Ali Z, Bithal PK, Prabhakar H, Rath GP, Dash HH. An assessment of the predictors of difficult intubation in patients with acromegaly. *Journal of Clinical Neuroscience* 2009;**16**(8):1043–5. PUBMED: 19457671]

Ali 2012 *{published data only}*
Ali MA, Qamar-ul-Hoda M, Samad K. Comparison of upper lip bite test with Mallampati test in the prediction of difficult intubation at a tertiary care hospital of Pakistan. *Journal of the Pakistan Medical Association* 2012;**62**(10):1012–5. PUBMED: 23866436]

Allahyary 2008 *{published data only}*
Allahyary E, Ghaemei SR, Azemati S. Comparison of six methods for predicting difficult intubation in obstetric patients. *Iranian Red Crescent Medical Journal* 2008;**10**(3):197–204.

Al Ramadhani 1996 *{published data only}*
Al Ramadhani S, Mohamed LA, Rocke DA, Gouws E. Sternomental distance as the sole predictor of difficult laryngoscopy in obstetric anaesthesia. *British Journal of Anaesthesia* 1996;**77**(3):312–6. PUBMED: 8949801]

Ambesh 2013 *{published data only}*
Ambesh SP, Singh N, Rao PB, Gupta D, Singh PK, Singh U. A combination of the modified Mallampati score, thyromental distance, anatomical abnormality, and cervical mobility (M-TAC) predicts difficult laryngoscopy better than Mallampati classification. *Acta Anaesthesiologica Taiwanica* 2013;**51**(2):58–62. PUBMED: 23968655]

Applegate 2013 *{published data only}*
Applegate RL 2nd, Gildea B, Patchin R, Rook JL, Wolford B, Nyirady J, et al. Telemedicine pre-anesthesia evaluation: a randomized pilot trial. *Telemedicine Journal and e-health* 2013;**19**(3):211–6. PUBMED: 23384334]

Arne 1998 *{published data only}*
Arné J, Descoins P, Fusciardi J, Ingrand P, Ferrier B, Boudigues D, et al. Preoperative assessment for difficult intubation in general and ENT surgery: predictive value of a clinical multivariate risk index. *British Journal of Anaesthesia* 1998;**80**(2):140. PUBMED: 9602574]

Ayhan 2016 *{published data only}*
Ayhan A, Kaplan S, Kayhan Z, Arslan G. Evaluation and management of difficult airway in obesity a single center retrospective study. *Acta Clinica Croatica* 2016;**55**(Suppl 1):27–32. PUBMED: 27276769]

Ayuso 2009 *{published data only}*
Ayuso MA, Sala X, Luis M, Carbó JM. Predicting difficult orotracheal intubation in pharyngo-laryngeal disease:

preliminary results of a composite index. *Canadian Journal of Anesthesia* 2003;**50**(1):81–5. PUBMED: 12514157]

Badheka 2016 ***{published data only}***

Badheka JP, Doshi PM, Vyas AM, Kacha NJ, Parmar VS. Comparison of upper lip bite test and ratio of height to thyromental distance with other airway assessment tests for predicting difficult endotracheal intubation. *Indian Journal of Critical Care Medicine* 2016;**20**(1):3–8. PUBMED: 26955210]

Baig 2014 ***{published data only}***

Baig MM, Khan FH. To compare the accuracy of prayer's sign and Mallampatti test in predicting difficult intubation in diabetic patients. *The Journal of the Pakistan Medical Association* 2014;**64**(8):879–83. PUBMED: 25252511]

Basaranoglu 2010 ***{published data only}***

Basaranoglu G, Columb M, Lyons G. Failure to predict difficult tracheal intubation for emergency caesarean section. *European Journal of Anaesthesiology* 2010;**27**(11): 947–9. PUBMED: 20864892]

Basunia 2013 ***{published data only}***

Basunia SR, Ghosh S, Bhattacharya S, Saha I, Biswas A, Prasad A. Comparison between different tests and their combination for prediction of difficult intubation: An analytical study. *Anesthesia, Essays and Researches* 2013;**7**(1): 105–9. PUBMED: 25885730]

Bergler 1997 ***{published data only}***

* Bergler W, Maleck W, Baker-Schreyer A, Petroianu G, Hörmann K. Difficult intubation in otorhinolaryngologic laser surgery. Is there a predictive parameter? [Schwierige Intubation in der HNO–Laserchirurgie Gibt es einen prädikativen Parameter?]. *Hals- Nasen- und Ohrenärzte* 1997;**45**(11):923–6. PUBMED: 9476105]
Bergler W, Maleck W, Baker-Schreyer A, Ungemach J, Petroianu G, Hörmann K. The Mallampati Score. Prediction of difficult intubation in otolaryngologic laser surgery by Mallampati Score [Der Mallampati–Score Vorhersage der schwierigen Intubation in der HNO–Laserchirurgie mittels Mallampati–Score]. *Anaesthesist* 1997;**46**(5):437–40. PUBMED: 9245215]

Bhat 2007 ***{published data only}***

Bhat RR, Mishra Sk, Badhe AS. Comparison of upper lip bite test and modified Mallampati classification in predicting difficult intubation. *The Internet Journal of Anesthesiology* 2007;**13**(1):1–4. https://print.ispub.com/api/0/ispub–article/7795]

Bilgin 1998 ***{published data only}***

Bilgin H, Ozyurt G. Screening tests for predicting difficult intubation. A clinical assessment in Turkish patients. *Anaesthesia and Intensive Care* 1998;**26**(4):382–6. PUBMED: 9743852]

Bindra 2010 ***{published data only}***

Adamus M. Comment on the article by Bindra A et al.: Is the modified Mallampati test performed in supine position a reliable predictor of difficult tracheal intubation?. *Journal of Anesthesia* 2011;**25**(1):135. PUBMED: 21052736]
* Bindra A, Prabhakar H, Singh GP, Ali Z, Singhal V. Is the modified Mallampati test performed in supine position a reliable predictor of difficult tracheal intubation?. *Journal of Anesthesia* 2010;**24**(3):482–5. PUBMED: 20217151]

Bouaggad 2004 ***{published and unpublished data}***

Bouaggad A, Nejmi SE, Bouderka MA, Abbassi O. Prediction of difficult tracheal intubation in thyroid surgery. *Anesthesia and Analgesia* 2004;**99**(2):603–6. PUBMED: 15271749]

Breckwoldt 2011 ***{published data only}***

Breckwoldt J, Klemstein S, Brunne B, Schnitzer L, Mochmann HC, Arntz HR. Difficult prehospital endotracheal intubation - predisposing factors in a physician based EMS. *Resuscitation* 2011;**82**(12):1519–24. PUBMED: 21749908]

Brodsky 2002 ***{published and unpublished data}***

Brodsky JB, Lemmens HJ, Brock-Utne JG, Vierra M, Saidman LJ. Morbid obesity and tracheal intubation. *Anesthesia and Analgesia* 2002;**94**(3):732–6. PUBMED: 11867407]

Butler 1992 ***{published data only}***

Butler PJ, Dhara SS. Prediction of difficult laryngoscopy: an assessment of the thyromental distance and Mallampati predictive tests. *Anaesthesia and Intensive Care* 1992;**20**(2): 139–42. PUBMED: 1595845]

Cattano 2004 ***{published data only}***

Cattano D, Panicucci E, Paolicchi A, Forfori F, Giunta F, Hagberg C. Risk factors assessment of the difficult airway: an Italian survey of 1956 patients. *Anesthesia and Analgesia* 2004;**99**(6):1774–9. PUBMED: 15562070]

Cattano 2014 ***{published data only}***

Cattano D, Katsiampoura A, Corso RM, Killoran PV, Cai C, Hagberg CA. Predictive factors for difficult mask ventilation in the obese surgical population. *F1000Research* 2014;**9**(3):239. PUBMED: 25485099]

Chaves 2009 ***{published data only}***

Chaves A, Carvalho S, Botelho M. Difficult endotracheal intubation In thyroid surgery: A retrospective study. *The Internet Journal of Anesthesiology* 2009;**22**(1):1–4.

Choi 2013 ***{published data only}***

Choi JW, Kim JA, Kim HK, Oh MS, Kim DK. Chest anteroposterior diameter affects difficulty of laryngoscopy for non-morbidly obese patients. *Journal of Anesthesia* 2013; **27**(4):563–8. PUBMED: 23455773]

Connor 2011 ***{published data only}***

Connor CW, Segal S. Accurate classification of difficult intubation by computerized facial analysis. *Anesthesia and Analgesia* 2011;**112**(1):84–93. PUBMED: 21081769]

Cortellazzi 2007 ***{published data only}***

Cortellazzi P, Minati L, Falcone C, Lamperti M, Caldiroli D. Predictive value of the El-Ganzouri multivariate risk index for difficult tracheal intubation: a comparison of Glidescope videolaryngoscopy and conventional Macintosh

laryngoscopy. *British Journal of Anaesthesia* 2007;**99**(6): 906–11. PUBMED: 17962241]

De Jong 2015 *{published data only}*

De Jong A, Molinari N, Pouzeratte Y, Verzilli D, Chanques G, Jung B, et al. Difficult intubation in obese patients: incidence, risk factors, and complications in the operating theatre and in intensive care units. *British Journal of Anaesthesia* 2015;**114**(2):297–306. PUBMED: 25431308]

Descoins 1994 *{published data only}*

Descoins P, Arné J, Bresard D, Ariès J, Fusciardi J. Proposal for a new multifactor screening score of difficult intubation in ORL and stomatognathic surgery: preliminary study. *Annales Françaises d'anesthèsie et de Rèanimation* 1994;**13**(2): 195–200. PUBMED: 7818203]

Dohrn 2015 *{published data only}*

Dohrn N, Sommer T, Bisgaard J, Rønholm E, Larsen JF. Difficult tracheal intubation in obese gastric bypass patients. *Obesity Surgery* 2016;**26**(11):2640–7. PUBMED: 26989060]

Domi 2009 *{published data only}*

Domi R. A comparison of Wilson Sum Score and Combination Mallampati, Tiromental and Sternomental Distances for predicting difficult intubation. *Macedonian Journal of Medical Sciences* 2009;**2**(2):141–4. DOI: 10.3889/MJMS.1857-5773.2009.0045

Domi 2010 *{published data only}*

Domi R. The best prediction test of difficult intubation. *Journal of Anaesthesiology Clinical Pharmacololgy* 2010;**26** (2):193–6.

Eberhart 2005 *{published data only}*

Eberhart LH, Arndt C, Cierpka T, Schwanekamp J, Wulf H, Putzke C. The reliability and validity of the upper lip bite test compared with the Mallampati classification to predict difficult laryngoscopy: an external prospective evaluation. *Anesthesia and Analgesia* 2005;**101**(1):284–9. PUBMED: 15976247]

El-Ganzouri 1996 *{published data only}*

el-Ganzouri AR, McCarthy RJ, Tuman KJ, Tanck EN, Ivankovich AD. Preoperative airway assessment: predictive value of a multivariate risk index. *Anesthesia and Analgesia* 1996;**82**(6):1197–204. PUBMED: 8638791]

Ezri 2001 *{published data only}*

Ezri T, Warters RD, Szmuk P, Saad-Eddin H, Geva D, Katz J, 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. *Anesthesia and Analgesia* 2001;**93**(4):1073–5. PUBMED: 11574386]

Ezri 2003a *{published data only}*

Ezri T, Medalion B, Weisenberg M, Szmuk P, Warters RD, Charuzi I. Increased body mass index per se is not a predictor of difficult laryngoscopy. *Canadian Journal of Anaesthesia = Journal Canadien d'anesthésie* 2003;**50**(2): 179–83. PUBMED: 12560311]

Ezri 2003b *{published data only}*

Ezri T, Weisenberg M, Khazin V, Zabeeda D, Sasson L, Shachner A, et al. Difficult laryngoscopy: incidence and predictors in patients undergoing coronary artery bypass surgery versus general surgery patients. *Journal of Cardiothoracic Anesthesia* 2003;**17**(3):321–4. PUBMED: 12827579]

Ezri 2003c *{published data only}*

Ezri T, Gewürtz G, Sessler DI, Medalion B, Szmuk P, Hagberg C, et al. Prediction of difficult laryngoscopy in obese patients by ultrasound quantification of anterior neck soft tissue. *Anaesthesia* 2003;**58**(11):1111–4. PUBMED: 14616599]

Frerk 1991 *{published data only}*

Frerk CM. Predicting difficult intubation. *Anaesthesia* 1991;**46**(12):1005–8. PUBMED: 1781521]

Frerk 1996 *{published data only}*

Frerk CM, Till CB, Bradley AJ. Difficult intubation: thyromental distance and the atlanto-occipital gap. *Anaesthesia* 1996;**51**(8):738–40. PUBMED: 8795315]

Freund 2012 *{published data only}*

Freund Y, Duchateau FX, Devaud ML, Ricard-Hibon A, Juvin P, Mantz J. Factors associated with difficult intubation in prehospital emergency medicine. *European Journal of Emergency Medicine* 2012;**19**(5):304–8. PUBMED: 22008586]

Fritscherova 2011 *{published data only}*

Fritscherova S, Adamus M, Dostalova K, Koutna J, Hrabalek L, Zapletalova J, et al. Can difficult intubation be easily and rapidly predicted?. *Biomedical Papers of the Medical Faculty of the University Palacký, Olomouc, Czechoslovakia* 2011;**155** (2):165–71. PUBMED: 21804626]

Gonzalez 2008 *{published data only}*

Gonzalez H, Minville V, Delanoue K, Mazerolles M, Concina D, Fourcade O. The importance of increased neck circumference to intubation difficulties in obese patients. *Anesthesia and Analgesia* 2008;**106**(4):1132–6. PUBMED: 18349184]

Hagberg 2009 *{published data only}*

Hagberg CA, Vogt-Harenkamp C, Kamal J. A retrospective analysis of airway management in obese patients at a teaching institution. *Journal of Clinical Anesthesia* 2009;**21** (5):348–51. PUBMED: 19700274]

Hagiwara 2015 *{published data only}*

Hagiwara Y, Watase H, Okamoto H, Goto T, Hasegawa K, Japanese Emergency Medicine Network Investigators. Prospective validation of the modified LEMON criteria to predict difficult intubation in the ED. *The American Journal of Emergency Medicine* 2015;**33**(10):1492–6. PUBMED: 26166379]

Hashim 2014 *{published data only}*

Hashim K, Thomas M. Sensitivity of palm print sign in prediction of difficult laryngoscopy in diabetes: A comparison with other airway indices. *Indian Journal of Anaesthesia* 2014;**58**(3):298–302. PUBMED: 25024473]

Healy 2016 *{published data only}*

Healy DW, LaHart EJ, Peoples EE, Jewell ES, Bettendorf RJ Jr, Ramachandran SK. A comparison of the Mallampati

evaluation in neutral or extended cervical spine positions: a retrospective observational study of > 80,000 patients. *British Journal of Anaesthesia* 2016;**116**(5):690–8. PUBMED: 27106973]

Heinrich 2013 ***{published data only}***

Heinrich S, Birkholz T, Irouschek A, Ackermann A, Schmidt J. Incidences and predictors of difficult laryngoscopy in adult patients undergoing general anesthesia: a single-center analysis of 102,305 cases. *Journal of Anesthesia* 2013;**27**(6): 815–21. PUBMED: 23748552]

Hekiert 2007 ***{published data only}***

Hekiert AM, Mandel J, Mirza N. Laryngoscopies in the obese: predicting problems and optimizing visualization. *Annals of Otology, Rhinology and Laryngology* 2007;**116**(4): 312–6. PUBMED: 17491533]

Hirmanpour 2014 ***{published data only}***

Hirmanpour A, Safavi M, Honarmand A, Jabalameli M, Banisadr G. The predictive value of the ratio of neck circumference to thyromental distance in comparison with four predictive tests for difficult laryngoscopy in obstetric patients scheduled for caesarean delivery. *Advanced Biomedical Research* 2014;**3**:200. PUBMED: 25337530]

Honarmand 2008 ***{published data only}***

Honarmand A, Safavi MR. Prediction of difficult laryngoscopy in obstetric patients scheduled for Caesarean delivery. *European Journal of Anaesthesiology* 2008;**25**(9): 714–20. PUBMED: 18471331]

Honarmand 2014 ***{published data only}***

Honarmand A, Safavi M, Ansari N. A comparison of between hyomental distance ratios, ratio of height to thyromental, modified Mallamapati classification test and upper lip bite test in predicting difficult laryngoscopy of patients undergoing general anesthesia. *Advanced Biomedical Research* 2014;**3**:166. PUBMED: 25221769]

Honarmand 2015 ***{published data only}***

Honarmand A, Safavi M, Yaraghi A, Attari M, Khazaei M, Zamani M. Comparison of five methods in predicting difficult laryngoscopy: Neck circumference, neck circumference to thyromental distance ratio, the ratio of height to thyromental distance, upper lip bite test and Mallampati test. *Advanced Biomedical Research* 2015;**4**:122. PUBMED: 26261824]

Huh 2009 ***{published data only}***

Huh J, Shin HY, Kim SH, Yoon TK, Kim DK. Diagnostic predictor of difficult laryngoscopy: the hyomental distance ratio. *Anesthesia and Analgesia* 2009;**108**(2):544–8. PUBMED: 19151285]

Hui 2009 ***{published data only}***

Hui C, Tsui BC. Sublingual ultrasound examination of the airway: a pilot study. *Canadian Journal of Anaesthesia = Journal Canadien d'anesthésie* 2009;**56**:S1–150.

Istvan 2010 ***{published data only}***

Istvan J, Belliveau M, Donati F. Rapid sequence induction for appendectomies: a retrospective case-review analysis. *Canadian Journal of Anaesthesia = Journal Canadien d'anesthésie* 2010;**57**(4):330–6. PUBMED: 20049576]

Ittichaikulthol 2010 ***{published data only}***

Ittichaikulthol W, Chanpradub S, Amnoundetchakorn S, Arayajarernwong N, Wongkum W. Modified Mallampati test and thyromental distance as a predictor of difficult laryngoscopy in Thai patients. *Journal of the Medical Association of Thailand = Chotmaihet Thangphaet* 2010;**93** (1):84–9. PUBMED: 20196416]

Juvin 2003 ***{published data only}***

Juvin P, Lavaut E, Dupont H, Lefevre P, Demetriou M, Dumoulin JL, et al. Difficult tracheal intubation is more common in obese than in lean patients. *Anesthesia and Analgesia* 2003;**97**(2):595–600. PUBMED: 12873960]

Kalezic 2016 ***{published data only}***

Kalezić N, Lakićević M, Miličić B, Stojanović M, Sabljak V, Marković D. Hyomental distance in the different head positions and hyomental distance ratio in predicting difficult intubation. *Bosnian Journal of Basic Medical Sciences / Udruženje Basičnih Mediciniskih Znanost* 2016;**16**(3):232–6. PUBMED: 27299374]

Kamalipour 2005 ***{published data only}***

Kamalipour H, Bagheri M, Kamali K, Taleie A, Yarmohammadi H. Lateral neck radiography for prediction of difficult orotracheal intubation. *European Journal of Anaesthesiology* 2005;**22**(9):689–93. PUBMED: 16163916]

Kamranmanesh 2013 ***{published data only}***

Kamranmanesh MR, Jafari AR, Gharaei B, Aghamohammadi H, Poor Zamany NKM, Kashi AH. Comparison of acromioaxillosuprasternal notch index (a new test) with modified Mallampati test in predicting difficult visualization of larynx. *Acta Anaesthesiologica Taiwanica* 2013;**51**(4):141–4. PUBMED: 24529668]

Khan 2003 ***{published data only}***

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. *Anesthesia and Analgesia* 2003;**96**(2):595–9. PUBMED: 12538218]

Khan 2009a ***{published data only}***

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 easy laryngoscopy and intubation: a prospective study. *Anesthesia and Analgesia* 2009;**109**(3):822–4. PUBMED: 19690252]

Khan 2009b ***{published data only}***

Khan ZH, Mofrad MK, Arbabi S, Javid MJ, Makarem J. Upper lip bite test as a predictor of difficult mask ventilation: a prospective study. *Middle East Journal of Anaesthesiology* 2009;**20**(3):377–82. PUBMED: 19950730]

Khan 2011 ***{published data only}***

Khan ZH, Maleki A, Makarem J, Mohammadi M, Khan RH, Zandieh A. A comparison of the upper lip bite test with hyomental/thyrosternal distances and mandible length

in predicting difficulty in intubation: A prospective study. *Indian Journal of Anaesthesia* 2011;**55**(1):43–6. PUBMED: 21431052]

Khan 2013 ***{published data only}***

Khan ZH, Arbabi S. Diagnostic value of the upper lip bite test in predicting difficulty in intubation with head and neck landmarks obtained from lateral neck X-ray. *Indian Journal of Anaesthesia* 2013;**57**(4):381–6. PUBMED: 24163453]

Khan 2014 ***{published data only}***

Khan ZH, Arbabi S, Yekaninejad MS, Khan RH. Application of the upper lip catch test for airway evaluation in edentulous patients: An observational study. *Saudi Journal of Anaesthesia* 2014;**8**(1):73–7. PUBMED: 24665244]

Khan 2015 ***{published data only}***

Khan ZH, Eskandari S, Yekaninejad MS. A comparison of the Mallampati test in supine and upright positions with and without phonation in predicting difficult laryngoscopy and intubation: A prospective study. *Journal of Anaesthesiology, Clinical Pharmacology* 2015;**31**(2):207–11. PUBMED: 25948902]

Kheterpal 2009 ***{published data only}***

Kheterpal S, Martin L, Shanks AM, Tremper KK. Prediction and outcomes of impossible mask ventilation: a review of 50,000 anesthetics. *Anesthesiology* 2009;**110**(4):891–7. PUBMED: 19293691]

Kim 2011 ***{published data only}***

Kim WH, Ahn HJ, Lee CJ, Shin BS, Ko JS, Choi SJ, et al. Neck circumference to thyromental distance ratio: a new predictor of difficult intubation in obese patients. *British Journal of Anaesthesia* 2011;**106**(5):743–8. PUBMED: 21354999]

K Nasa 2014 ***{published data only}***

K Nasa V, S Kamath S. Risk factors assessment of the difficult intubation using intubation difficulty scale (IDS). *Journal of Clinical and Diagnostic Research: JCDR* 2014;**8**(7):GC01–3. PUBMED: 25177576]

Knudsen 2014 ***{published data only}***

Knudsen K, Högman M, Larsson A, Nilsson U. The best method to predict easy intubation: a quasi-experimental pilot study. *Journal of Perianesthesia Nursing* 2014;**29**(4):292–7. PUBMED: 25062573]

Koh 2002 ***{published data only}***

Koh LK, Kong CE, Ip-Yam PC. The modified Cormack-Lehane score for the grading of direct laryngoscopy: evaluation in the Asian population. *Anaesthesia and Intensive Care* 2002;**30**(1):48–51. PUBMED: 11939440]

Kolarkar 2015 ***{published data only}***

Kolarkar P, Badwaik G, Watve A, Abhishek K, Bhangale N, Bhalerao A, et al. Upper lip bite test: A novel test of predicting difficulty in intubation. *Journal of Evolution of Medical and Dental Sciences* 2015;**4**(24):4149–56.

Komatsu 2007 ***{published data only}***

Komatsu R, Sengupta P, Wadhwa A, Akça O, Sessler DI, Ezri T, et al. Ultrasound quantification of anterior soft tissue thickness fails to predict difficult laryngoscopy in obese patients. *Anaesthesia and Intensive Care* 2007;**35**(1):32–7. PUBMED: 17323663]

Konwar 2015 ***{published data only}***

Konwar C, Baruah ND, Saikia P, Chakrabartty A. A prospective study of the usefulness of upper lip bite test in combination with sternomental distance, thyromental distance and inter-incisor distance as predictor of ease of laryngoscopy. *Journal of Evolution of Medical and Dental Sciences* 2015;**4**(97):16286–89.

Krobbuaban 2005 ***{published data only}***

Krobbuaban B, Diregpoke S, Kumkeaw S. An assessment of the ratio of height to thyromental distance compared to thyromental distance as a predictive test for prediction of difficult tracheal intubation in Thai patients. *Journal of the Medical Association of Thailand = Chotmaihet Thangphaet* 2006;**89**(5):638–42. PUBMED: 16756049]

* Krobbuaban B, Diregpoke S, Kumkeaw S, Tanomsat M. The predictive value of the height ratio and thyromental distance: four predictive tests for difficult laryngoscopy. *Anesthesia and Analgesia* 2005;**101**(5):1542–5. PUBMED: 16244029]

Langeron 2000 ***{published data only}***

Langeron O, Masso E, Huraux C, Guggiari M, Bianchi A, Coriat P, et al. Prediction of difficult mask ventilation. *Anesthesiology* 2000;**92**(5):1229–36. PUBMED: 10781266]

Lee 2015 ***{published data only}***

Lee SL, Hosford C, Lee QT, Parnes SM, Shapshay SM. Mallampati class, obesity, and a novel airway trajectory measurement to predict difficult laryngoscopy. *Laryngoscope* 2015;**125**(1):161–6. PUBMED: 25042696]

Liaskou 2014 ***{published data only}***

Liaskou C, Vouzounerakis E, Moirasgenti M, Trikoupi A, Staikou C. Anatomic features of the neck as predictive markers of difficult direct laryngoscopy in men and women: A prospective study. *Indian Journal of Anaesthesia* 2014;**58**(2):176–82. PUBMED: 24963183]

Lundstrom 2009 ***{published data only}***

* Lundstrøm LH, Møller AM, Rosenstock C, Astrup G, Gätke MR, Wetterslev J, et al. A documented previous difficult tracheal intubation as a prognostic test for a subsequent difficult tracheal intubation in adults. *Anaesthesia* 2009;**64**(10):1081–8. PUBMED: 19735399]

Lundstrøm LH, Møller AM, Rosenstock C, Astrup G, Gätke MR, Wetterslev J, et al. Avoidance of neuromuscular blocking agents may increase the risk of difficult tracheal intubation: a cohort study of 103,812 consecutive adult patients recorded in the Danish Anaesthesia Database. *British Journal of Anaesthesia* 2009;**103**(2):283–90. PUBMED: 19457894]

Lundstrøm LH, Møller AM, Rosenstock C, Astrup G, Wetterslev J. High body mass index is a weak predictor for difficult and failed tracheal intubation: a cohort study of 91,332 consecutive patients scheduled for direct laryngoscopy registered in the Danish Anesthesia Database. *Anesthesiology* 2009;**110**(2):266–74. PUBMED: 19194154]

Mallat 2010 *{published data only}*
Mallat J, Robin E, Pironkov A, Lebuffe G, Tavernier B. Goitre and difficulty of tracheal intubation. *Annales Françaises d'anesthèsie et de Rèanimation* 2010;**296**(6): 436–9. PUBMED: 20547033]

Mashour 2008 *{published data only}*
Mashour GA, Kheterpal S, Vanaharam V, Shanks A, Wang LY, Sandberg WS, et al. The extended Mallampati score and a diagnosis of diabetes mellitus are predictors of difficult laryngoscopy in the morbidly obese. *Anesthesia and Analgesia* 2008;**107**(6):1919–23. PUBMED: 19020139]

Mehta 2014 *{published data only}*
Mehta T, Jayaprakash J, Shah V. Diagnostic value of different screening tests in isolation or combination for predicting difficult intubation: A prospective study. *Indian Journal of Anaesthesia* 2014;**58**(6):754–7. PUBMED: 25624545]

Merah 2004 *{published data only}*
* Merah NA, Foulkes-Crabbe DJ, Kushimo OT, Ajayi PA. Prediction of difficult laryngoscopy in a population of Nigerian obstetric patients. *West African Journal of Medicine* 2004;**23**(1):38–41. PUBMED: 15171524]
Merah NA, Wong DT, Ffoulkes-Crabbe DJ, Kushimo OT, Bode CO. Modified Mallampati test, thyromental distance and inter-incisor gap are the best predictors of difficult laryngoscopy in West Africans. *Canadian Journal of Anesthesia* 2005;**52**(3):291–6. PUBMED: 15753502]

Mishra 2009 *{published data only}*
Mishra SK, Bhat RR, Sudeep K, Nagappa M, Badhe AS. Comparison of upper lip bite test with modified Mallampati classification for prediction of difficult obstetric intubation. *The Internet Journal of Anesthesiology* 2010;**19**(1):1–3.

Montemayor-Cruz 2015 *{published data only}*
Montemayor-Cruz JM, Guerrero-Ledezma RM. Diagnostic utility of the hyomental distance ratio as predictor of difficult intubation at UMAE 25. *Gaceta Médica de México* 2015;**151**(5):599–607. PUBMED: 26526473]

Myneni 2010 *{published data only}*
Myneni N, O'Leary AM, Sandison M, Roberts K. Evaluation of the upper lip bite test in predicting difficult laryngoscopy. *Journal of Clinical Anesthesia* 2010;**22**(3): 174–8. PUBMED: 20400002]

Nadal 1998 *{published data only}*
* Nadal JL, Fernandez BG, Escobar IC, Black M, Rosenblatt WH. The palm print as a sensitive predictor of difficult laryngoscopy in diabetics. *Acta Anaesthesiologica Scandinavica* 1998;**42**(2):199–203. PUBMED: 9509203]
Rae R, Kinsella J, Daniel M, Booth M. The palm print as predictor of difficult intubation. *Acta Anaesthesiologica Scandinavica* 2001;**45**(8):1053. PUBMED: 11576065]

Naguib 1999 *{published data only}*
Naguib M, Malabarey T, AlSatli RA, Al Damegh S, Samarkandi AH. Predictive models for difficult laryngoscopy and intubation. A clinical, radiologic and three-dimensional computer imaging study. *Canadian Journal of Anaesthesia = Journal Canadien d'anesthésie* 1999; **46**(8):748–59. PUBMED: 10451134]

Naguib 2006 *{published data only}*
Naguib M, Scamman FL, O'Sullivan C, Aker J, Ross AF, Kosmach S, et al. Predictive performance of three multivariate difficult tracheal intubation models: a double-blind, case-controlled study. *Anesthesia and Analgesia* 2006; **102**(3):818–24. PUBMED: 16492835]

Nasir 2011 *{published data only}*
Nasir KK, Shahani AS, Maqbool MS. Correlative value of airway assessment by Mallampati classification and Cormack and Lehane grading. *Rawal Medical Journal* 2011; **36**:2–4.

Nasiri 2013 *{published data only}*
Nasiri E, Akbari H, Farrokhi H, Ebrahimi F, Nasiri R. A comparison of the Upper Lip Bite Test with Inter Incisor Distance in predicting difficult laryngoscopy and tracheal intubation. *Journal of Mazandaran University of Medical Sciences* 2013;**23**(105):71–9.

Nath 1997 *{published data only}*
Nath G, Sekar M. Predicting difficult intubation - a comprehensive scoring system. *Anaesthesia and Intensive Care* 1997;**25**(5):482–6. PUBMED: 9352759]

Noorizad 2006 *{published data only}*
Noorizad S, Mahdian M. Mallampati and thyromental distance to predict difficult intubation. *Journal of Medical Sciences* 2006;**6**(2):169–72.

Oates 1990 *{published data only}*
Oates JD, Oates PD, Pearsall FJ, McLeod AD, Howie JC. Phonation affects Mallampati class. *Anaesthesia* 1990;**45** (11):984. PUBMED: 2082969]

Oates 1991 *{published data only}*
Oates JD, Macleod AD, Oates PD, Pearsall FJ, Howie JC, Murray GD. Comparison of two methods for predicting difficult intubation. *British Journal of Anaesthesia* 1991;**66** (3):305–9. PUBMED: 2015145]

Pottecher 1991 *{published data only}*
Pottecher T, Velten M, Galani M, Forrler M. Comparative value of clinical signs of difficult tracheal intubation in women. *Annales Françaises d'anesthèsie et de Rèanimation* 1991;**10**(5):430–5. PUBMED: 1755552]

Prakash 2013 *{published data only}*
Prakash S, Kumar A, Bhandari S, Mullick P, Singh R, Gogia AR. Difficult laryngoscopy and intubation in the Indian population: An assessment of anatomical and clinical risk factors. *Indian Journal of Anaesthesia* 2013;**57**(6):569–75. PUBMED: 24403616]

Qudaisat 2011 *{published data only}*
Qudaisat IY, Al-Ghanem SM. Short thyromental distance is a surrogate for inadequate head extension, rather than small submandibular space, when indicating possible difficult direct laryngoscopy. *European Journal of Anaesthesiology* 2011;**28**(8):600–6. PUBMED: 21610502]

Reghunathan 2016 *{published data only}*
Reghunathan U, Raghavan RK, Veetil BM. Airway assessment for anticipation of difficult intubation: a double

blind comparative study. *Journal of Evolution Medical and Dental Sciences* 2016;**5**(40):2466–9.

Rocke 1992 ***{published data only}***
Rocke DA, Murray WB, Rout CC, Gouws E. Relative risk analysis of factors associated with difficult intubation in obstetric anesthesia. *Anesthesiology* 1992;**77**(1):67–73. PUBMED: 1610011]

Safavi 2014 ***{published data only}***
Safavi M, Honarmand A, Amoushahi M. Prediction of difficult laryngoscopy: Extended mallampati score versus the MMT, ULBT and RHTMD. *Advances in Biomedical Engineering Research* 2014;**3**:133. PUBMED: 24949304]

Sahin 2011 ***{published data only}***
Sahin SH, Yilmaz A, Gunday I, Kargi M, Sut N, Taskinalp O, et al. Using temporomandibular joint mobility to predict difficult tracheal intubation. *Journal of Anesthesia* 2011;**25**(3):457–61. PUBMED: 21451975]

Salimi 2008 ***{published data only}***
Salimi A, Farzanegan B, Rastegarpour A, Kolahi AA. Comparison of the upper lip bite test with measurement of thyromental distance for prediction of difficult intubations. *Acta Anaesthesiologica Taiwanica* 2008;**46**(2):61–5. PUBMED: 18593650]

Samra 1995 ***{published data only}***
Samra SK, Schork MA, Guinto FC Jr. A study of radiologic imaging techniques and airway grading to predict a difficult endotracheal intubation. *Journal of Clinical Anesthesia* 1995;**7**(5):373–9. PUBMED: 7576671]

Savva 1994 ***{published data only}***
Savva D. Prediction of difficult tracheal intubation. *British Journal of Anaesthesia* 1994;**73**(2):149–53. PUBMED: 7917726]

Schmitt 2000 ***{published data only}***
Schmitt H, Buchfelder M, Radespiel-Tröger M, Fahlbusch R. Difficult intubation in acromegalic patients: incidence and predictability. *Anesthesiology* 2000;**93**(1):110–4. PUBMED: 10861153]

Seo 2012 ***{published data only}***
Seo SH, Lee JG, Yu SB, Kim DS, Ryu SJ, Kim KH. Predictors of difficult intubation defined by the intubation difficulty scale (IDS): predictive value of 7 airway assessment factors. *Korean Journal of Anesthesiology* 2012;**63**(6):491–7. PUBMED: 23277808]

Shah 2013 ***{published data only}***
Shah PJ, Dubey KP, Yadav JP. Predictive value of upper lip bite test and ratio of height to thyromental distance compared to other multivariate airway assessment tests for difficult laryngoscopy in apparently normal patients. *Journal of Anaesthesiology, Clinical Pharmacology* 2013;**29**(2):191–5. PUBMED: 23878440]

Shah 2014 ***{published data only}***
Shah AA, Rafique K, Islam M. Can difficult intubation be accurately predicted using upper lip bite test?. *Journal of Postgraduate Medical Institute* 2014;**28**(3):282–7.

Sharma 2010 ***{published data only}***
Sharma D, Prabhakar H, Bithal PK, Ali Z, Singh GP, Rath GP, et al. Predicting difficult laryngoscopy in acromegaly: a comparison of upper lip bite test with modified Mallampati classification. *Journal of Neurosurgical Anesthesiology* 2010;**22**(2):138–43. PUBMED: 20118795]

Singh 2009 ***{published data only}***
Singh R, Jain A, Swatantra M, Kohli P. Clinical evaluation of predicting difficult laryngoscopy in obstetric patients. *Journal of Anaesthesiology, Clinical Pharmacology* 2009;**25**(1):38–42.

Soyuncu 2009 ***{published data only}***
Soyuncu S, Eken C, Cete Y, Bektas F, Akcimen M. Determination of difficult intubation in the ED. *The American Journal of Emergency Medicine* 2009;**27**(8):905–10. PUBMED: 19857405]

Tantri 2016 ***{published data only}***
Tantri AR, Firdaus R, Salomo ST. Predictors of difficult intubation among Malay patients in Indonesia. *Anesthesiology and Pain Medicine* 2016;**6**(2):e34848. PUBMED: 27252905]

Thompson 2009 ***{published data only}***
Thompson J, O'Neill SS, Hutchings L, Jones R. Retrospective study of 1602 obstetric intubations: predicting difficult and failed intubation using the Mallampati test. *International Journal of Obstetric Anesthesia* 2009;**18**:S1–63. DOI: https://doi.org/10.1016/j.ijoa.2009.03.006

Tse 1995 ***{published data only}***
Tse JC, Rimm EB, Hussain A. Predicting difficult endotracheal intubation in surgical patients scheduled for general anesthesia: a prospective blind study. *Anesthesia and Analgesia* 1995;**81**(2):254–8. PUBMED: 7618711]

Tuzuner-Oncul 2008 ***{published data only}***
Tuzuner-Oncul AM, Kucukyavuz Z. Prevalence and prediction of difficult intubation in maxillofacial surgery patients. *Journal of Oral and Maxillofacial Surgery* 2008;**66**(8):1652–8. PUBMED: 18634954]

Ul Haq 2013 ***{published data only}***
Ul Haq MI, Ullah H. Comparison of Mallampati test with lower jaw protrusion maneuver in predicting difficult laryngoscopy and intubation. *Journal of Anaesthesiology, Clinical Pharmacology* 2013;**29**(3):313–7. PUBMED: 24106353]

Uribe 2015 ***{published data only}***
Uribe AA, Zvara DA, Puente EG, Otey AJ, Zhang J, Bergese SD. BMI as a predictor for potential difficult tracheal intubation in males. *Frontiers in Medicine (Lausanne)* 2015;**2**(38):e. PUBMED: 26137460]

Vallem 2015 ***{published data only}***
Vallem B, Thalisetty J, Challapalli SR, Israel N, Gudise S, Murthigari S. Comparison of upper lip bite test with other four predictors for predicting difficulty in intubation. *Journal of Evolution of Medical and Dental Sciences* 2015;**4**(39):6811–7. DOI: 10.14260/jemds/2015/988

Vani 2000 *{published data only}*
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. *Journal of Postgraduate Medicine* 2000;**46**(2):75–9. PUBMED: 11013469]

Wajekar 2015 *{published data only}*
Wajekar AS, Chellam S, Toal PV. Prediction of ease of laryngoscopy and intubation-role of upper lip bite test, modified mallampati classification, and thyromental distance in various combination. *Journal of Family Medicine and Primary Care* 2015;**4**(1):101–5. PUBMED: 25810998]

Wilson 1988 *{published data only}*
Wilson ME, Spiegelhalter D, Robertson JA, Lesser P. Predicting difficult intubation. *British Journal of Anaesthesia* 1988;**61**(2):211–6. PUBMED: 3415893]

Wong 1999 *{published data only}*
Wong SH, Hung CT. Prevalence and prediction of difficult intubation in Chinese women. *Anaesthesia and Intensive Care* 1999;**27**(1):49–52. PUBMED: 10050224]

Wong 2009 *{published data only}*
Wong P, Parrington S. Difficult intubation in ENT and maxillofacial surgical patients: a prospective survey. *The Internet Journal of Anesthesiology* 2009;**21**(1):1–3.

Yamamoto 1997 *{published data only}*
Yamamoto K, Tsubokawa T, Shibata K, Ohmura S, Nitta S, Kobayashi T. Predicting difficult intubation with indirect laryngoscopy. *Anesthesiology* 1997;**86**(2):316–21. PUBMED: 9054250]

Yildiz 2005 *{published data only}*
Yildiz TS, Solak M, Toker K. The incidence and risk factors of difficult mask ventilation. *Journal of Anesthesia* 2005;**19**(1):7–11. PUBMED: 15674508]

Yildiz 2007 *{published data only}*
Yildiz TS, Korkmaz F, Solak M, Toker K, Erciyes N, Bayrak F, et al. Prediction of difficult tracheal intubation in Turkish patients: a multi-center methodological study. *European Journal of Anaesthesiology* 2007;**24**(12):1034–40. PUBMED: 17555609]

Yu 2015 *{published data only}*
Yu T, Wang B, Jin XJ, Wu RR, Wu H, He JJ, et al. Predicting difficult airways: 3-3-2 rule or 3-3 rule?. *Irish Journal of Medical Science* 2015;**184**(3):677–83. PUBMED: 25740093]

References to studies excluded from this review

Acer 2011 *{published data only}*
Acer N, Akkaya A, Tugay BU, Öztürk A. A comparison of Cormack-Lehane and Mallampati tests with mandibular and neck measurements for predicting difficult intubation [Zor Entübasyonu Tahmin Etmek İçin Cormeck–Lehane ve Mallampati Testleri İle Mandibula ve Boyun Ölçümlerinin Karşılaştırılmaları]. *Balkan Medical Journal* 2011;**28**:157–63. DOI: 10.5174/tutfd.2010.03475.1

Acikgoz 2015 *{published data only}*
Acikgoz AO, Karagoz H, Yilbas AA, Akca B, Uzumcugil F, Pamuk G. Difficult airway and risk factors in bariatric surgery patients. *Bariatric Surgical Practice and Patient Care* 2015;**10**(4):145–9. DOI: 10.1089/bari.2015.0026

Beyus 2010 *{published data only}*
Beyus C, Mort T. Challenges of airway management in obesity. Thirty-Ninth Critical Care Congress of the Society of Critical Care Medicine; 2010 Jan 9-13; Miami Beach (FL). Miami Beach, FL: Society of Critical Care Medicine, 2010.

Hiremath 1998 *{published data only}*
Hiremath AS, Hillman DR, James AL, Noffsinger WJ, Platt PR, Singer SL. Relationship between difficult tracheal intubation and obstructive sleep apnoea. *British Journal of Anaesthesia* 1998;**80**:606–11.

Lewis 1994 *{published data only}*
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?. *Anaesthesiology* 1994;**81**:69–75.

Meininger 2010 *{published data only}*
Meininger D, Strouhal U, Weber CF, Fogl D, Holzer L, Zacharowski K, et al. Direct laryngoscopy or C-MAC video laryngoscopy? Routine tracheal intubation in patients undergoing ENT surgery [Direkte laryngoskopie oder C–MAC– Vvdeolaryngoskopie? Routineintubation von patienten in der HNO–heilkunde]. *Anaesthesist* 2010;**59**:806–11.

Moon 2013 *{published data only}*
Moon HY, Baek CW, Kim JS, Koo GH, Kim JY, Woo YC, et al. The causes of difficult tracheal intubation and preoperative assessments in different age groups. *Korean Journal of Anesthesiology* 2013;**64**(4):308–14. PUBMED: 23646239]

Oriol-López 2009 *{published data only}*
Oriol-Lopez SA, Hernandez-Bernal CE. Assessment, prediction and presence of difficult intubation [Valoración, predicción y presencia de intubación difícil]. *Revista Mexicana de Anestesiologica* 2009;**32**(1):41–9.

Orozco-Díaz 2010 *{published data only}*
Orozco-Díaz E, Alvarez-Ríos JJ, Arceo-Díaz JL, Ornelas-Aguirre JM. Predictive factors of difficult airway with known assessment scales. *Cirugia y Cirujanos* 2010;**78**(5):393–9. PUBMED: 21219809]

Reed 2005 *{published data only}*
Reed MJ, Dunn MJ, McKeown DW. Can an airway assessment score predict difficulty at intubation in the emergency department?. *Emergency Medicine Journal* 2005;**22**(2):99–102. PUBMED: 15662057]

Safavi 2011 *{published data only}*
Safavi M, Honarmand A, Zare N. A comparison of the ratio of patient's height to thyromental distance with the modified

Mallampati and the upper lip bite test in predicting difficult laryngoscopy. *Saudi Journal of Anaesthesia* 2011;**5**(3): 258–63. PUBMED: 21957403]

Siyam 2002 ***{published data only}***
Siyam MA, Benhamou D. Difficult endotracheal intubation in patients with sleep apnea syndrome. *Anesthesia and Analgesia* 2002;**95**(4):1098–102. PUBMED: 12351303]

Tripathi 2006 ***{published data only}***
Tripathi M, Pandey M. Short thyromental distance: a predictor of difficult intubation or an indicator for small blade selection?. *Anesthesiology* 2006;**104**(6):1131–6. PUBMED: 16732082]

References to studies awaiting assessment

Akhlaghi 2017 ***{published data only}***
Akhlaghi M, Abedinzadeh M, Ahmadi A, Heidari Z. Predicting difficult laryngoscopy and intubation with laryngoscopic exam test: a new method. *Acta Medica Iranica* 2017;**55**:453–8.

Andrade 2017 ***{published data only}***
Andrade R, Lima BL, Lopes DK, Couceiro Filho RO, Lima LC, Couceiro TC. Difficult laryngoscopy and tracheal intubation: observational study. *Revista Brasileira de Anestesiologia* 2017;**20**:20.

Awan 2017 ***{published data only}***
Awan A, Rehman A. Comparison of neck circumference to thyromental distance ratio with modified Mallampati score for prediction of difficult intubation in obese patients. *Acta Anaesthesiologica Scandinavica* 2017;**61**:997.

Banik 2017 ***{published data only}***
Banik D, Ray L, Akhtaruzzaman AK, Bhowmick DK, Hossain MS, Islam MS, et al. Assessment of difficulties associated with endotracheal intubation using modified mallampati and upper lip bite test. *Mymensingh Medical Journal: MMJ* 2017;**26**:395–405.

Belda 2017 ***{published data only}***
Belda I, Ayuso MA, Sala-Blanch X, Luis M, Berge R. A predictive test for difficult intubation in laryngeal microsurgery. Validation study. *Revista Espanola de Anestesiologia y Reanimacion* 2017;**64**:71–8.

Card 2017 ***{published data only}***
Card ME, Rucci J, Honiden S, Heavner J. Identifying factors associated with difficult airway during endotracheal intubation in the medical ICU. *Chest* 2017;**152**(4 Suppl 1): A218.

Carlson 2017 ***{published data only}***
Carlson JN, Hostler D, Guyette FX, Pinchalk M, Martin-Gill C. Derivation and validation of the prehospital difficult airway identification tool (PreDAIT): a predictive model for difficult intubation. *The Western Journal of Emergency Medicine* 2017;**18**:662–72.

Dar 2017 ***{published data only}***
Dar S, Khan MS, Iqbal F, Nazeer T, Hussain R. Comparison of upper lip bite test (ULBT) with mallampati classification, regarding assessment of difficult intubation. *Pakistan Journal of Medical and Health Sciences* 2017;**11**:767–9.

Eiamcharoenwit 2017 ***{published data only}***
Eiamcharoenwit J, Itthisompaiboon N, Limpawattana P, Suwanpratheep A, Siriussawakul A. The performance of neck circumference and other airway assessment tests for the prediction of difficult intubation in obese parturients undergoing cesarean delivery. *International Journal of Obstetric Anesthesia* 2017;**31**:45–50.

Han 2017 ***{published data only}***
Han YZ, Tian Y, Xu M, Ni C, Li M, Wang J, et al. Neck circumference to inter-incisor gap ratio: A new predictor of difficult laryngoscopy in cervical spondylosis patients. *BMC Anesthesiology* 2017;**17**(1):1–3.

Hanouz 2018 ***{published data only}***
Hanouz JL, Bonnet V, Buleon C, Simonet T, Radenac D, Zamparini G, et al. Comparison of the mallampati classification in sitting and supine position to predict difficult tracheal intubation: a prospective observational cohort study. *Anesthesia and Analgesia* 2018;**126**:161–9.

Jain 2017 ***{published data only}***
Jain N, Das S, Kanchi M. Thyromental height test for prediction of difficult laryngoscopy in patients undergoing coronary artery bypass graft surgical procedure. *Annals of Cardiac Anaesthesia* 2017;**20**:207–11.

Khatiwada 2017 ***{published data only}***
Khatiwada S, Bhattarai B, Pokharel K, Acharya R. Prediction of difficult airway among patients requiring endotracheal intubation in a tertiary care hospital in eastern Nepal. *JNMA; Journal of the Nepal Medical Association* 2017;**56** (207):314–8.

Lee 2017 ***{published data only}***
Lee SY, Chien DK, Huang MY, Huang CH, Shih SC, Wu KM, et al. Patient-specific factors associated with difficult mask ventilation in the emergency department. *International Journal of Gerontology* 2017;**11**:263–6.

Mahmoodpoor 2017 ***{published data only}***
Mahmoodpoor A, Soleimanpour H, Golzari SE, Nejabatian A, Pourlak T, Amani M, et al. Determination of the diagnostic value of the modified mallampati score, upper lip bite test and facial angle in predicting difficult intubation: a prospective descriptive study. *Journal of Clinical Anesthesia* 2017;**37**:99–102.

Norskov 2017 ***{published data only}***
Norskov AK, Wetterslev J, Rosenstock CV, Afshari A, Astrup G, Jakobsen JC, et al. Prediction of difficult mask ventilation using a systematic assessment of risk factors vs. existing practice - a cluster randomised clinical trial in 94, 006 patients. Anaesthesia 2017; Vol. 72, issue 3:296–308.

Prakash 2017 ***{published data only}***
Prakash S, Mullick P, Bhandari S, Kumar A, Gogia AR, Singh R. Sternomental distance and sternomental displacement as predictors of difficult laryngoscopy and intubation in adult patients. *Saudi Journal of Anaesthesia* 2017;**11**:273–8.

Rao 2017 *{published data only}*
Rao CS, Ranganath T, Rao SP, Sujani K. Comparison of upper lip bite test with modified mallampati test and thyromental distance for predicting difficulty in endotracheal intubation - a prospective study. *Journal of Evolution of Medical and Dental Sciences* 2017;**6**:1413–6.

Riad 2018 *{published data only}*
Riad W, Ansari T, Shetty N. Does neck circumference help to predict difficult intubation in obstetric patients? A prospective observational study. *Saudi Journal of Anaesthesia* 2018;**12**:77–81.

Selvi 2017 *{published data only}*
Selvi O, Kahraman T, Senturk O, Tulgar S, Serifsoy E, Ozer Z. Evaluation of the reliability of preoperative descriptive airway assessment tests in prediction of the Cormack-Lehane score: a prospective randomized clinical study. Journal of Clinical Anesthesia 2017; Vol. 36:21–6.

Shankar 2017 *{published data only}*
Shankar D, Suresh YV. Comparison of various airway assessment factors with ratio of height to thyromental distance (rhtmd) in predicting difficult airway in apparently normal patients. *Journal of Evolution of Medical and Dental Sciences* 2017;**6**:902–7.

Siljeblad 2017 *{published data only}*
Siljeblad M, Lofgren B, Snygg J, Nellgard P. A comparison of predictive factors for identifying difficult airway (laryngoscopy & intubation) patients. *Acta Anaesthesiologica Scandinavica* 2017;**61**(8):1036–7.

Srivilaithon 2018 *{published data only}*
Srivilaithon W, Muengtaweepongsa S, Sittichanbuncha Y, Patumanond J. Predicting difficult intubation in emergency department by intubation assessment score. *Journal of Clinical Medicine Research* 2018;**10**:247–53.

Torres 2017 *{published data only}*
Torres K, Blonski M, Pietrzyk L, Piasecka-Twarog M, Maciejewski R, Torres A. Usefulness and diagnostic value of the NEMA parameter combined with other selected bedside tests for prediction of difficult intubation. *Journal of Clinical Anesthesia* 2017;**37**:132–5.

Wang 2017 *{published data only}*
Wang B, Yao W D, Peng H, Guo L, Jin X J. Interincisor distance predicting difficulty airway in men and women. *Anesthesia and Analgesia* 2017;**124**:23–4.

Workeneh 2017 *{published data only}*
Workeneh SA, Gebregzi AH, Denu ZA. Magnitude and predisposing factors of difficult airway during induction of general anaesthesia. *Anesthesiology Research and Practice* 2017;**2017**:1–6.

Yildirim 2017 *{published data only}*
Yildirim I, Inal MT, Memis D, Turan FN. Determining the efficiency of different preoperative difficult intubation tests on patients undergoing caesarean section. *Balkan Medical Journal* 2017;**34**:436–43.

Additional references

Adnet 1997
Adnet F, Borron SW, Racine SX, Clemessy JL, Fournier JL, Plaisance P, et al. The intubation difficulty scale (IDS): proposal and evaluation of a new score characterizing the complexity of endotracheal intubation. *Anesthesiology* 1997; **87**(6):1290–7. [PUBMED: 9416711]

ASA 2003
American Society of Anesthesiologists. 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**(5):1269–77. [PUBMED: 12717151]

ASA 2013
Apfelbaum JL, Hagberg CA, Caplan RA, Blitt CD, Connis RT, Nickinovich DG, et al. American Society of Anesthesiologists Task Force on Management of the Difficult Airway. 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* 2013;**118**(2):251–70. [PUBMED: 23364566]

Begg 2005
Begg CB. Systematic reviews of diagnostic accuracy studies require study by study examination: first for heterogeneity, and then for sources of heterogeneity. *Journal of Clinical Epidemiology* 2005;**58**(9):865–6. [PUBMED: 16085189]

Berkow 2009
Berkow LC, Greenberg RS, Kan KH, Colantuoni E, Mark LJ, Flint PW, et al. Need for emergency surgical airway reduced by a comprehensive difficult airway program. *Anesthesia and Analgesia* 2009;**109**(6):1860–9. [PUBMED: 19713264]

Calder 2003
Calder I, Picard J, Chapman M, O'Sullivan C, Crockard HA. Mouth opening: a new angle. *Anesthesiology* 2003;**99** (4):799–801. [PUBMED: 14508309]

Chu 2006
Chu H, Cole SR. Bivariate meta-analysis of sensitivity and specificity with sparse data: a generalized linear mixed model approach. *Journal of Clinical Epidemiology* 2006;**59** (2):1331–2.

Combes 2006
Combes X, Jabre P, Jbeili C, Leroux B, Bastuji-Garin S, Margenet A, et al. Prehospital standardization of medical airway management: incidence and risk factors of difficult airway. *Academic Emergency Medicine* 2006;**13**(8):828–34. [PUBMED: 16807397]

Cook 2000
Cook TM. A new practical classification of laryngeal view. *Anaesthesia* 2000;**55**(3):274–9. [PUBMED: 10671848]

Cormack 1984
Cormack RS, Lehane J. Difficult tracheal intubation in obstetrics. *Anaesthesia* 1984;**39**(11):1105–11. [PUBMED: 6507827]

DA Society 2015
Frerk C, Mitchell VS, McNarry AF, Mendonca C, Bhagrath R, Patel A, et al. Difficult Airway Society 2015 guidelines for management of unanticipated difficult intubation in adults. *British Journal of Anaesthesia* 2015;**115**(6):827–48. [PUBMED: 26556848]

El-Orbany 2009
El-Orbany M, Woehlck HJ. Difficult mask ventilation. *Anesthesia and Analgesia* 2009;**109**(6):1870–80. [PUBMED: 19923516]

Karkouti 1996
Karkouti K, Rose DK, Ferris LE, Wigglesworth DF, Meisami-Fard T, Lee H. Inter-observer reliability of ten tests used for predicting difficult tracheal intubation. *Canadian Journal of Anaesthesia* 1996;**43**(6):554–9. [PUBMED: 8773859]

Lee 2006
Lee A, Fan LT, Gin T, Karmakar MK, Ngan Kee WD. A systematic review (meta-analysis) of the accuracy of the Mallampati tests to predict the difficult airway. *Anesthesia and Analgesia* 2006;**102**(6):1867–78. [PUBMED: 16717341]

Liberati 2009
Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gotzsche PC, Ioannidis JP, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *PLoS Medicine* 2009;**6**(6):e1000100.

Luba 2010
Luba K, Cutter TW. Supraglottic airway devices in the ambulatory setting. *Anesthesiology Clinics* 2010;**28**: 295–314. [PUBMED: 20488396]

Lundstrom 2011
Lundstrøm LH, Vester-Andersen M, Møller AM, Charuluxananan S, L'hermite J, Wetterslev J, et al. Poor prognostic value of the modified Mallampati score: a meta-analysis involving 177 088 patients. *British Journal of Anaesthesia* 2011;**107**(5):659–67. [PUBMED: 21948956]

Mallampati 1985
Mallampati SR, Gatt SP, Gugino LD, Desai SP, Waraksa B, Freiberger D, et al. A clinical sign to predict difficult tracheal intubation: a prospective study. *Canadian Anaesthetists' Society Journal* 1985;**32**(4):429–34. [PUBMED: 4027773]

Partlett 2016
Partlett C, Takwoingi Y. Meta-analysis of test accuracy studies in R: a summary of user-written programs and step-by-step guide to using glmer. Version 1.0. www.methods.cochrane.org/sdt/ (accessed 2 August 2016).

Peterson 2005
Peterson GN, Domino KB, Caplan RA, Posner KL, Lee LA, Cheney FW. Management of the difficult airway: a closed claims analysis. *Anesthesiology* 2005;**103**(1):33–9. [PUBMED: 15983454]

Pott 2008
Pott LM, Murray WB. Review of video laryngoscopy and rigid fiberoptic laryngoscopy. *Current Opinion in Anaesthesiology* 2008;**21**(6):750–8. [PUBMED: 18997526]

R 2017 [Computer program]
R Foundation for Statistical Computing. R: A language and environment for statistical computing. Version 3.4.2. Vienna, Austria: R Foundation for Statistical Computing, 2017.

Ramadhani 1996
Ramadhani SAL, Mohamed LA, Rocke DA, Gouws E. Sternomental distance as the sole predictor of difficult laryngoscopy in obstetric anaesthesia. *British Journal of Anaesthesia* 1996;**77**(3):312–6. [PUBMED: 8949801]

Review Manager 2014 [Computer program]
Nordic Cochrane Centre, The Cochrane Collaboration. Review Manager 5 (RevMan 5). Version 5.3. Copenhagen: Nordic Cochrane Centre, The Cochrane Collaboration, 2014.

Rose 1996
Rose DK, Cohen MM. The incidence of airway problems depends on the definition used. *Canadian Journal of Anaesthesia* 1996;**43**(1):30–4. [PUBMED: 8665631]

Samsoon 1987
Samsoon GL, Young JR. Difficult tracheal intubation: a retrospective study. *Anaesthesia* 1987;**42**(5):487–90. [PUBMED: 3592174]

Shiga 2005
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):429–37. [PUBMED: 16052126]

Stata 2015 [Computer program]
StataCorp. Stata. Version 14. College Station, TX, USA: StataCorp, 2015.

Whiting 2003
Whiting P, Rutjes AW, Reitsma JB, Bossuyt PM, Kleijnen J. The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. *BMC Medical Research Methodology* 2003;**3**:25. [PUBMED: 14606960]

Whiting 2011
Whiting PF, Rutjes AWS, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, et al. QUADAS-2 Group. QUADAS-2: A revised tool for the quality assessment of diagnostic accuracy studies. *Annals of Internal Medicine* 2011;**155**(8):529–36. [PUBMED: 22007046]

Yentis 1998
Yentis SM, Lee DJ. Evaluation of an improved scoring system for the grading of direct laryngoscopy. *Anaesthesia* 1998;**53**(11):1041–4. [PUBMED: 10023271]

References to other published versions of this review

Lee 2010

Lee A, Herkner H, Hovhannisyan K, Pace NL. Airway physical examination tests for detection of difficult airway management in apparently normal patients. *Cochrane Database of Systematic Reviews* 2010, Issue 12. DOI: 10.1002/14651858.CD008874

* *Indicates the major publication for the study*

CHARACTERISTICS OF STUDIES

Characteristics of included studies *[ordered by study ID]*

Adamus 2010

Study characteristics			
Patient sampling	All adult patients selected for endotracheal intubation for non-emergency surgical procedures		
Patient characteristics and setting	Sample size: 1518 763 females		
Index tests	Modified Mallampati		
Target condition and reference standard(s)	Difficult laryngoscopy: following induction of general anaesthesia and muscle relaxation, direct laryngoscopy was performed. The laryngoscopic view under optimal conditions ("morning air sniffing position") was described		
Flow and timing	Index test: on arrival at operating theatre Reference standard: following induction of general anaesthesia		
Comparative			
Notes			
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?			
Was a case-control design avoided?			
Did the study avoid inappropriate exclusions?			
DOMAIN 2: Index Test All Tests			
Were the index test results interpreted without knowledge of the results of the reference standard?			

If a threshold was used, was it pre-specified?			
Did the assessors of the index test have appropriate training?			
Was interobserver variability reported for some or all patients?			
Was interobserver agreement acceptable?			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?			
Were the reference standard results interpreted without knowledge of the results of the index tests?			
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?			
Did all patients receive the same reference standard?			
Were all patients included in the analysis?			

Adnet 2001

Study characteristics	
Patient sampling	All consecutive surgical patients scheduled for anaesthesia requiring tracheal intubation were studied

Adnet 2001 (*Continued*)

Patient characteristics and setting	University hospital, surgical patients (abdominal, cardiac, thoracic, orthopaedic, ENT surgery) Sample size: 1171 505 females Mean age: 49 years		
Index tests	MMT, ULBT		
Target condition and reference standard(s)	Difficult laryngoscopy, IDS > 5		
Flow and timing	Preoperative visit to surgery		
Comparative			
Notes			
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?			
Was a case-control design avoided?			
Did the study avoid inappropriate exclusions?			
DOMAIN 2: Index Test All Tests			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Did the assessors of the index test have appropriate training?			
Was interobserver variability reported for some or all patients?			

Adnet 2001 (*Continued*)

Was interobserver agreement acceptable?			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?			
Were the reference standard results interpreted without knowledge of the results of the index tests?			
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?			
Did all patients receive the same reference standard?			
Were all patients included in the analysis?			

Aktas 2015

Study characteristics	
Patient sampling	"Selected at random" excluding emergency operations, patients needing awake intubation and patients with congenital anomalies
Patient characteristics and setting	Sample size: 120 67 females Mean age: 47.5 years
Index tests	MMT
Target condition and reference standard(s)	Difficult laryngoscopy: Cormack and Lehane

Aktas 2015 (*Continued*)

Flow and timing	Not stated		
Comparative			
Notes			
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?			
Was a case-control design avoided?			
Did the study avoid inappropriate exclusions?			
DOMAIN 2: Index Test All Tests			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Did the assessors of the index test have appropriate training?			
Was interobserver variability reported for some or all patients?			
Was interobserver agreement acceptable?			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?			

Aktas 2015 (*Continued*)

Were the reference standard results interpreted without knowledge of the results of the index tests?			
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?			
Did all patients receive the same reference standard?			
Were all patients included in the analysis?			

Al Ramadhani 1996

Study characteristics			
Patient sampling	Patients planned for caesarian section were evaluated. If patients were selected for general anaesthesia, they were included		
Patient characteristics and setting	Sample size: 523 523 females Mean age: 30.4 years		
Index tests	SMD (13.5 cm)		
Target condition and reference standard(s)	Difficult laryngoscopy: Cormack and Lehane		
Flow and timing	Index test during pre-anaesthesia assesment Target condition after RSI		
Comparative			
Notes			
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**

Al Ramadhani 1996 (*Continued*)

DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?			
Was a case-control design avoided?			
Did the study avoid inappropriate exclusions?			
DOMAIN 2: Index Test All Tests			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Did the assessors of the index test have appropriate training?			
Was interobserver variability reported for some or all patients?			
Was interobserver agreement acceptable?			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?			
Were the reference standard results interpreted without knowledge of the results of the index tests?			
DOMAIN 4: Flow and Timing			

Al Ramadhani 1996 (*Continued*)

Was there an appropriate interval between index test and reference standard?			
Did all patients receive the same reference standard?			
Were all patients included in the analysis?			

Ali 2009

Study characteristics			
Patient sampling	Over 5 years, 66 consecutive patients with acromegaly who presented for pituitary surgery		
Patient characteristics and setting	Consecutive patients with acromegaly who presented for pituitary surgery Sample size: 66 32 females Mean age: 43.4 years		
Index tests	MMT		
Target condition and reference standard(s)	Difficult laryngoscopy: Cormack and Lehane		
Flow and timing	The time interval was not described. Modified Mallampati grade was assessed preoperatively		
Comparative			
Notes			
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?			
Was a case-control design avoided?			
Did the study avoid inappropriate exclusions?			

DOMAIN 2: Index Test All Tests			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Did the assessors of the index test have appropriate training?			
Was interobserver variability reported for some or all patients?			
Was interobserver agreement acceptable?			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?			
Were the reference standard results interpreted without knowledge of the results of the index tests?			
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?			
Did all patients receive the same reference standard?			
Were all patients included in the analysis?			

Ali 2012

Study characteristics			
Patient sampling	Adult patients (> 18 years of age) undergoing elective surgeries requiring general anaesthesia with endotracheal intubations were enrolled		
Patient characteristics and setting	Edentulous patients, those unable to open the mouth or with limited cervical movement or requiring rapid sequence induction were excluded Sample size: 324 199 females Mean age: 43 years		
Index tests	MMT, ULBT		
Target condition and reference standard(s)	Difficult laryngoscopy: laryngoscopy was done with Macintosh laryngoscope blade size 3 or 4, and laryngoscopic view of the first attempt at intubation was graded and recorded according to Cormack and Lehane classification with the patient in the sniffing position but without applying external laryngeal pressure		
Flow and timing	Preoperatively; no further information		
Comparative			
Notes			
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?			
Was a case-control design avoided?			
Did the study avoid inappropriate exclusions?			
DOMAIN 2: Index Test All Tests			
Were the index test results interpreted without knowledge of the results of the reference standard?			

If a threshold was used, was it pre-specified?			
Did the assessors of the index test have appropriate training?			
Was interobserver variability reported for some or all patients?			
Was interobserver agreement acceptable?			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?			
Were the reference standard results interpreted without knowledge of the results of the index tests?			
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?			
Did all patients receive the same reference standard?			
Were all patients included in the analysis?			

Allahyary 2008

Study characteristics	
Patient sampling	Consecutive women were prospectively included Exclusion criteria were gross anatomical abnormality or recent surgery of the head and neck, preeclampsia, severe cardiorespiratory disorders, inability to sit and edentulous patients

Allahyary 2008 (*Continued*)

Patient characteristics and setting	Consecutive obstetric parturients with ASA I/II undergoing general anaesthesia for caesarean delivery Sample size: 203 203 females		
Index tests	MMT, TMD, SMD (13.5cm), mouth opening, ULBT, combination of tests		
Target condition and reference standard(s)	Difficult laryngoscopy: Cormack and Lehane		
Flow and timing	No time interval defined, but the tests were performed preoperatively		
Comparative			
Notes			
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?			
Was a case-control design avoided?			
Did the study avoid inappropriate exclusions?			
DOMAIN 2: Index Test All Tests			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Did the assessors of the index test have appropriate training?			
Was interobserver variability reported for some or all patients?			

Allahyary 2008 (*Continued*)

Was interobserver agreement acceptable?			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?			
Were the reference standard results interpreted without knowledge of the results of the index tests?			
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?			
Did all patients receive the same reference standard?			
Were all patients included in the analysis?			

Ambesh 2013

Study characteristics	
Patient sampling	Consecutive adult patients ASA I and II undergoing scheduled general anaesthesia
Patient characteristics and setting	Obvious difficult airway excluded Sample size: 500 208 females Mean age: 46 years
Index tests	MMT, combination of tests
Target condition and reference standard(s)	Difficult larnygoscopy: Cormack and Lehane, Macintosh blade after general anesthesia and muscle relaxation

Ambesh 2013 (*Continued*)

Flow and timing	Preoperatively; no further information		
Comparative			
Notes			
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?			
Was a case-control design avoided?			
Did the study avoid inappropriate exclusions?			
DOMAIN 2: Index Test All Tests			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Did the assessors of the index test have appropriate training?			
Was interobserver variability reported for some or all patients?			
Was interobserver agreement acceptable?			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?			

Ambesh 2013 (*Continued*)

Were the reference standard results interpreted without knowledge of the results of the index tests?			
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?			
Did all patients receive the same reference standard?			
Were all patients included in the analysis?			

Applegate 2013

Study characteristics			
Patient sampling	Adult patients scheduled for head and neck surgery with the ability to read, write, and speak English were considered for inclusion		
Patient characteristics and setting	Sample size: 160 70 females Mean age: 55 years		
Index tests	Combination of tests		
Target condition and reference standard(s)	Difficult laryngoscopy: Cormack and Lehane, Macintosh blade after general anesthesia and muscle relaxation		
Flow and timing	Preoperatively; no further information		
Comparative			
Notes			
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			

Was a consecutive or random sample of patients enrolled?			
Was a case-control design avoided?			
Did the study avoid inappropriate exclusions?			
DOMAIN 2: Index Test All Tests			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Did the assessors of the index test have appropriate training?			
Was interobserver variability reported for some or all patients?			
Was interobserver agreement acceptable?			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?			
Were the reference standard results interpreted without knowledge of the results of the index tests?			
DOMAIN 4: Flow and Timing			

Applegate 2013 (*Continued*)

Was there an appropriate interval between index test and reference standard?			
Did all patients receive the same reference standard?			
Were all patients included in the analysis?			

Arne 1998

Study characteristics			
Patient sampling	During an 18-month period, any patient older than 15 years of age undergoing ENT or general surgery with tracheal intubation was considered as potentially eligible		
Patient characteristics and setting	Sample size: 1200 Mean age: 47 years		
Index tests	MMT, TMD, mouth opening (< 5 cm)		
Target condition and reference standard(s)	DIfficult tracheal intubation: patient placed in optimal (sniffing) position on OR table Anaesthesia induced followed by neuromuscular blockade Macintosh blade at first attempt		
Flow and timing	Index tests during preoperative consultation		
Comparative			
Notes			
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?			
Was a case-control design avoided?			
Did the study avoid inappropriate exclusions?			

DOMAIN 2: Index Test All Tests			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Did the assessors of the index test have appropriate training?			
Was interobserver variability reported for some or all patients?			
Was interobserver agreement acceptable?			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?			
Were the reference standard results interpreted without knowledge of the results of the index tests?			
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?			
Did all patients receive the same reference standard?			
Were all patients included in the analysis?			

Ayhan 2016

Study characteristics			
Patient sampling	Adult patients surgically treated for endometrial cancer from January 2011 to December 2014		
Patient characteristics and setting	Patients operated for endometrial cancer were reviewed and only those patients with BMI ≥ 25 (N = 427) were included in the study Sample size: 427 427 females Mean age: 58 years		
Index tests	MMT, mouth opening		
Target condition and reference standard(s)	DIfficult laryngoscopy: Cormack and Lehane; difficult face mask ventilation		
Flow and timing	During preoperative visit and during induction of anaesthesia		
Comparative			
Notes			
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?			
Was a case-control design avoided?			
Did the study avoid inappropriate exclusions?			
DOMAIN 2: Index Test All Tests			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			

Ayhan 2016 (*Continued*)

Did the assessors of the index test have appropriate training?			
Was interobserver variability reported for some or all patients?			
Was interobserver agreement acceptable?			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?			
Were the reference standard results interpreted without knowledge of the results of the index tests?			
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?			
Did all patients receive the same reference standard?			
Were all patients included in the analysis?			

Ayuso 2009

Study characteristics	
Patient sampling	Consecutive adult patients with laryngeal disease to undergo laryngeal microsurgry under general anaesthesia
Patient characteristics and setting	No prior testing Sample size: 181 47 females

Ayuso 2009 (*Continued*)

	Mean age: 53.5 years		
Index tests	MMT, TMD, mouth opening (< 4 cm)		
Target condition and reference standard(s)	Difficult laryngoscopy: Cormack and Lehane		
Flow and timing	Not stated		
Comparative			
Notes			
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?			
Was a case-control design avoided?			
Did the study avoid inappropriate exclusions?			
DOMAIN 2: Index Test All Tests			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Did the assessors of the index test have appropriate training?			
Was interobserver variability reported for some or all patients?			
Was interobserver agreement acceptable?			

Ayuso 2009 (*Continued*)

DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?			
Were the reference standard results interpreted without knowledge of the results of the index tests?			
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?			
Did all patients receive the same reference standard?			
Were all patients included in the analysis?			

Badheka 2016

Study characteristics	
Patient sampling	Patients of both gender between 20 and 70 years of age with ASA I-III scheduled to undergo elective surgery under general anaesthesia with endotracheal intubation
Patient characteristics and setting	Patients with airway malformation, oral surgery, neck burns contracture, midline neck swelling, emergency surgery, caesarean section, edentulous patients, limitation of temporomandibular/atlantoaxial joint mobility, and history of neck surgery were excluded from the study Sample size: 170 73 females
Index tests	MMT, TMD (< 6 cm), SMD, mouth opening, ULBT
Target condition and reference standard(s)	Difficult laryngoscopy: laryngoscopy was done by a qualified and experienced anaesthesiologist, who was blinded to the results of preoperative airway assessment and glottic visualization were assessed and noted according to modified Cormack and Lehane grade

Badheka 2016 (*Continued*)

Flow and timing	Preoperatively		
Comparative			
Notes			
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?			
Was a case-control design avoided?			
Did the study avoid inappropriate exclusions?			
DOMAIN 2: Index Test All Tests			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Did the assessors of the index test have appropriate training?			
Was interobserver variability reported for some or all patients?			
Was interobserver agreement acceptable?			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?			

Badheka 2016 (*Continued*)

Were the reference standard results interpreted without knowledge of the results of the index tests?			
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?			
Did all patients receive the same reference standard?			
Were all patients included in the analysis?			

Baig 2014

Study characteristics			
Patient sampling	Patients having ASA II status, age above 18 years and known cases of diabetes mellitus planned for elective surgeries requiring general anaesthesia		
Patient characteristics and setting	Those who already had airway deformity due to surgical or medical problem or those undergoing rapid sequence induction were excluded Sample size: 357 145 females Mean age: 53.7 years		
Index tests	MMT, combination of tests		
Target condition and reference standard(s)	Difficult laryngoscopy: Cormack and Lehane		
Flow and timing	Preoperatively		
Comparative			
Notes			
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**

DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?			
Was a case-control design avoided?			
Did the study avoid inappropriate exclusions?			
DOMAIN 2: Index Test All Tests			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Did the assessors of the index test have appropriate training?			
Was interobserver variability reported for some or all patients?			
Was interobserver agreement acceptable?			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?			
Were the reference standard results interpreted without knowledge of the results of the index tests?			
DOMAIN 4: Flow and Timing			

Baig 2014 (*Continued*)

Was there an appropriate interval between index test and reference standard?			
Did all patients receive the same reference standard?			
Were all patients included in the analysis?			

Basaranoglu 2010

Study characteristics			
Patient sampling	Consecutive patients for emergency caesarean delivery		
Patient characteristics and setting	No prior testing, routine evaluation Sample size: 239 239 females Mean age: 28 years		
Index tests	MMT, TMD (< 6 cm), SMD, mouth opening (< 3 cm), combination of tests		
Target condition and reference standard(s)	Difficult laryngoscopy: Cormack and Lehane		
Flow and timing	Within Minutes		
Comparative			
Notes			
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?			
Was a case-control design avoided?			
Did the study avoid inappropriate exclusions?			

DOMAIN 2: Index Test All Tests			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Did the assessors of the index test have appropriate training?			
Was interobserver variability reported for some or all patients?			
Was interobserver agreement acceptable?			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?			
Were the reference standard results interpreted without knowledge of the results of the index tests?			
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?			
Did all patients receive the same reference standard?			
Were all patients included in the analysis?			

Basunia 2013

Study characteristics			
Patient sampling	Patients (16 to 60 years), ASA I and II, scheduled for elective surgical procedures requiring ETI were included		
Patient characteristics and setting	Patients with inoral growth, unable to open mouth, chin on chest challenged person, pregnancy, previous history of difficult intubation, acquired disorders of head and neck were excluded Sample size: 300		
Index tests	MMT, SMD, combination of tests		
Target condition and reference standard(s)	Difficult laryngoscopy: Cormack and Lehane		
Flow and timing	Preoperatively		
Comparative			
Notes			
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?			
Was a case-control design avoided?			
Did the study avoid inappropriate exclusions?			
DOMAIN 2: Index Test All Tests			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Did the assessors of the index test have appropriate training?			

Basunia 2013 (*Continued*)

Was interobserver variability reported for some or all patients?			
Was interobserver agreement acceptable?			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?			
Were the reference standard results interpreted without knowledge of the results of the index tests?			
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?			
Did all patients receive the same reference standard?			
Were all patients included in the analysis?			

Bergler 1997

Study characteristics	
Patient sampling	Patients with planned laser surgery (ENT) Patients with reduced mobility were excluded
Patient characteristics and setting	Sample size: 91 22 females Mean age: 54 years
Index tests	MMT

Bergler 1997 (*Continued*)

Target condition and reference standard(s)	Difficult laryngoscopy: Cormack and Lehane		
Flow and timing	No details given		
Comparative			
Notes			
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?			
Was a case-control design avoided?			
Did the study avoid inappropriate exclusions?			
DOMAIN 2: Index Test All Tests			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Did the assessors of the index test have appropriate training?			
Was interobserver variability reported for some or all patients?			
Was interobserver agreement acceptable?			
DOMAIN 3: Reference Standard			

Bergler 1997 (*Continued*)

Is the reference standards likely to correctly classify the target condition?			
Were the reference standard results interpreted without knowledge of the results of the index tests?			
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?			
Did all patients receive the same reference standard?			
Were all patients included in the analysis?			

Bhat 2007

Study characteristics	
Patient sampling	ASA 1/2 patients, admitted for elective surgical procedure were prospectively included
Patient characteristics and setting	Excluded: edentulous patients, restricted mouth opening, restricted cervical movement Presence of oropharyngeal, laryngeal pathology Sample size: 500 286 females
Index tests	MMT, ULBT, combination of tests
Target condition and reference standard(s)	Difficult laryngoscopy: Cormack and Lehane
Flow and timing	Not described
Comparative	
Notes	
Methodological quality	

Bhat 2007 (*Continued*)

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?			
Was a case-control design avoided?			
Did the study avoid inappropriate exclusions?			
DOMAIN 2: Index Test All Tests			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Did the assessors of the index test have appropriate training?			
Was interobserver variability reported for some or all patients?			
Was interobserver agreement acceptable?			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?			
Were the reference standard results interpreted without knowledge of the results of the index tests?			

Bhat 2007 (*Continued*)

DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?			
Did all patients receive the same reference standard?			
Were all patients included in the analysis?			

Bilgin 1998

Study characteristics			
Patient sampling	ASA I-II for GA requiring endotracheal intubation. Excluded known abnormalities of upper airway or head and neck trauma		
Patient characteristics and setting	Sample size: 500 253 females Mean age: 45.6 years		
Index tests	MT, TMD (< 6 cm)		
Target condition and reference standard(s)	Difficult laryngoscopy: using Macintosh blade size 3, with head in "sniffing" position on a pillow. Cormack and Lehane III/IV defined as difficult. Difficult tracheal intubation		
Flow and timing	Exact timing not specified		
Comparative			
Notes			
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?			
Was a case-control design avoided?			

Bilgin 1998 (*Continued*)

Did the study avoid inappropriate exclusions?			
DOMAIN 2: Index Test All Tests			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Did the assessors of the index test have appropriate training?			
Was interobserver variability reported for some or all patients?			
Was interobserver agreement acceptable?			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?			
Were the reference standard results interpreted without knowledge of the results of the index tests?			
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?			
Did all patients receive the same reference standard?			

Bilgin 1998 (*Continued*)

Were all patients included in the analysis?			

Bindra 2010

Study characteristics			
Patient sampling	Not stated		
Patient characteristics and setting	No prior test; standard assessment; routine preoperative care Sample size: 123 52 females Mean age: 38 years		
Index tests	MMT		
Target condition and reference standard(s)	Difficult laryngoscopy: Cormack and Lehane		
Flow and timing	Not stated, but apparently just prior to being taken to operating theatre		
Comparative			
Notes			
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?			
Was a case-control design avoided?			
Did the study avoid inappropriate exclusions?			
DOMAIN 2: Index Test All Tests			

Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Did the assessors of the index test have appropriate training?			
Was interobserver variability reported for some or all patients?			
Was interobserver agreement acceptable?			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?			
Were the reference standard results interpreted without knowledge of the results of the index tests?			
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?			
Did all patients receive the same reference standard?			
Were all patients included in the analysis?			

Bouaggad 2004

Study characteristics			
Patient sampling	All patients aged 18 years and older scheduled to undergo thyroid surgery under general anaesthesia were prospectively included in the study. Patients with obvious malformations of the airway were excluded from the study		
Patient characteristics and setting	Patients undergoing elective thyroid surgery Sample size: 320 281 females		
Index tests	MMT, TMD (< 6 cm), mouth opening		
Target condition and reference standard(s)	Difficult laryngoscopy: the laryngeal view was assessed with rigid laryngoscopy by a certified anaesthesiologist or certified nurse anaesthetist using a Macintosh laryngoscope, Blade 3 or 4. Difficult tracheal intubation: evaluated by IDS		
Flow and timing	Unclear		
Comparative			
Notes			
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?			
Was a case-control design avoided?			
Did the study avoid inappropriate exclusions?			
DOMAIN 2: Index Test All Tests			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			

Bouaggad 2004 (*Continued*)

Did the assessors of the index test have appropriate training?			
Was interobserver variability reported for some or all patients?			
Was interobserver agreement acceptable?			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?			
Were the reference standard results interpreted without knowledge of the results of the index tests?			
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?			
Did all patients receive the same reference standard?			
Were all patients included in the analysis?			

Breckwoldt 2011

Study characteristics	
Patient sampling	All ETIs performed by the emergency physicians of the mobile intensive care unit and the helicopter emergency medical system were included
Patient characteristics and setting	Sample size: 276 105 females Mean age: 65 years

Breckwoldt 2011 (*Continued*)

Index tests	Mouth opening		
Target condition and reference standard(s)	Difficult laryngoscopy: Cormack and Lehane		
Flow and timing	Directly before intubation		
Comparative			
Notes			
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?			
Was a case-control design avoided?			
Did the study avoid inappropriate exclusions?			
DOMAIN 2: Index Test All Tests			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Did the assessors of the index test have appropriate training?			
Was interobserver variability reported for some or all patients?			
Was interobserver agreement acceptable?			

Breckwoldt 2011 (*Continued*)

DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?			
Were the reference standard results interpreted without knowledge of the results of the index tests?			
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?			
Did all patients receive the same reference standard?			
Were all patients included in the analysis?			

Brodsky 2002

Study characteristics	
Patient sampling	Consecutive
Patient characteristics and setting	Morbidly obese patients (BMI > 40) undergoing elective surgery Sample size: 100 78 females Mean age: 44 years
Index tests	MMT
Target condition and reference standard(s)	Difficult laryngoscopy: Cormack and Lehane; difficult tracheal intubation
Flow and timing	Unclear
Comparative	
Notes	

Brodsky 2002 (*Continued*)

Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?			
Was a case-control design avoided?			
Did the study avoid inappropriate exclusions?			
DOMAIN 2: Index Test All Tests			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Did the assessors of the index test have appropriate training?			
Was interobserver variability reported for some or all patients?			
Was interobserver agreement acceptable?			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?			
Were the reference standard results interpreted without knowledge of the results of the index tests?			

Brodsky 2002 (*Continued*)

DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?			
Did all patients receive the same reference standard?			
Were all patients included in the analysis?			

Butler 1992

Study characteristics			
Patient sampling	Mixed surgical patients		
Patient characteristics and setting	Patients with known abnormalities of the airway or with head or neck trauma were excluded Sample size: 250 153 females		
Index tests	MMT, TMD (< 6 cm)		
Target condition and reference standard(s)	Difficult laryngoscopy: Cormack and Lehane		
Flow and timing	No information		
Comparative			
Notes			
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?			
Was a case-control design avoided?			

Did the study avoid inappropriate exclusions?			
DOMAIN 2: Index Test All Tests			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Did the assessors of the index test have appropriate training?			
Was interobserver variability reported for some or all patients?			
Was interobserver agreement acceptable?			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?			
Were the reference standard results interpreted without knowledge of the results of the index tests?			
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?			
Did all patients receive the same reference standard?			

Butler 1992 (*Continued*)

Were all patients included in the analysis?			

Cattano 2004

Study characteristics			
Patient sampling	Consecutive		
Patient characteristics and setting	Adult patients scheduled to receive general anaesthesia requiring endotracheal intubation for elective abdominal, vascular, urologic, and endocrinologic surgery Sample size: 1956		
Index tests	MMT, TMD, SMD (< 1.2 cm), mouth opening, combination of tests		
Target condition and reference standard(s)	Difficult laryngoscopy: Cormack and Lehane; difficult face mask ventilation		
Flow and timing	Preoperatively		
Comparative			
Notes			
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?			
Was a case-control design avoided?			
Did the study avoid inappropriate exclusions?			
DOMAIN 2: Index Test All Tests			
Were the index test results interpreted without knowledge of the results of the reference stan-			

dard?			
If a threshold was used, was it pre-specified?			
Did the assessors of the index test have appropriate training?			
Was interobserver variability reported for some or all patients?			
Was interobserver agreement acceptable?			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?			
Were the reference standard results interpreted without knowledge of the results of the index tests?			
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?			
Did all patients receive the same reference standard?			
Were all patients included in the analysis?			

Cattano 2014

Study characteristics			
Patient sampling	A retrospective investigation was performed. 1399 anaesthetics were identified where both mask ventilation was attempted and a preprocedure airway evaluation was documented. Of these, 557 obese patients were identified and included for analysis		
Patient characteristics and setting	Obese patients Sample size: 557 307 females		
Index tests	MMT		
Target condition and reference standard(s)	Difficult face mask ventilation		
Flow and timing	Unknown		
Comparative			
Notes			
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?			
Was a case-control design avoided?			
Did the study avoid inappropriate exclusions?			
DOMAIN 2: Index Test All Tests			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Did the assessors of the index test have appropriate training?			

Was interobserver variability reported for some or all patients?			
Was interobserver agreement acceptable?			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?			
Were the reference standard results interpreted without knowledge of the results of the index tests?			
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?			
Did all patients receive the same reference standard?			
Were all patients included in the analysis?			

Chaves 2009

Study characteristics	
Patient sampling	Chart review
Patient characteristics and setting	Patients having elective thyroid surgery between January 2005 and June 2007; routine anaesthesia care About 10% of patients had clinical signs of tracheal compression or radiographic signs of intrathoracic goiter or tracheal compression in cervical radiogram Sample size: 512 448 females Mean age: 55 years

Chaves 2009 (*Continued*)

Index tests	MMT, TMD (< 6 cm), mouth opening (< 5 cm), combination of tests		
Target condition and reference standard(s)	Difficult tracheal intubation: more than three attempts necessary or a change in materials used		
Flow and timing	Not given		
Comparative			
Notes			
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?			
Was a case-control design avoided?			
Did the study avoid inappropriate exclusions?			
DOMAIN 2: Index Test All Tests			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Did the assessors of the index test have appropriate training?			
Was interobserver variability reported for some or all patients?			
Was interobserver agreement acceptable?			

Chaves 2009 (*Continued*)

DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?			
Were the reference standard results interpreted without knowledge of the results of the index tests?			
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?			
Did all patients receive the same reference standard?			
Were all patients included in the analysis?			

Choi 2013

Study characteristics	
Patient sampling	Consecutive patients of ASA I or II, aged 18 to 70 years old, and who were scheduled to undergo elective surgery under general anaesthesia were considered for enrolment
Patient characteristics and setting	Patients with loose upper incisors, airway pathology, gross anatomical abnormalities, BMI more than 35 kg/m2, or any history of difficult intubation were excluded Sample size: 269 132 females
Index tests	MMT
Target condition and reference standard(s)	Difficult laryngoscopy: Cormack and Lehane
Flow and timing	Preoperatively
Comparative	

Notes			
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?			
Was a case-control design avoided?			
Did the study avoid inappropriate exclusions?			
DOMAIN 2: Index Test All Tests			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Did the assessors of the index test have appropriate training?			
Was interobserver variability reported for some or all patients?			
Was interobserver agreement acceptable?			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?			
Were the reference standard results interpreted without knowledge			

Choi 2013 (*Continued*)

of the results of the index tests?			
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?			
Did all patients receive the same reference standard?			
Were all patients included in the analysis?			

Connor 2011

Study characteristics			
Patient sampling	"Patients meeting our entry criteria were identified by examination of their anesthesia records in the postanesthesia care unit"		
Patient characteristics and setting	Only one sex, one ethnicity Sample size: 80 0 female		
Index tests	MMT, TMD		
Target condition and reference standard(s)	Difficult laryngoscopy: Cormack and Lehane		
Flow and timing	Standard preoperative examination		
Comparative			
Notes			
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?			

Was a case-control design avoided?			
Did the study avoid inappropriate exclusions?			
DOMAIN 2: Index Test All Tests			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Did the assessors of the index test have appropriate training?			
Was interobserver variability reported for some or all patients?			
Was interobserver agreement acceptable?			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?			
Were the reference standard results interpreted without knowledge of the results of the index tests?			
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?			

Did all patients receive the same reference standard?			
Were all patients included in the analysis?			

Cortellazzi 2007

Study characteristics			
Patient sampling	"Data of consecutive patients intubated in two 5-month periods in 2004 and 2006 were anonymously entered in a computerized database for the purpose of the present study"		
Patient characteristics and setting	Sample size: 1837 885 females		
Index tests	Combination of tests		
Target condition and reference standard(s)	Difficult laryngoscopy: Cormack and Lehane		
Flow and timing	Index test was performed 1 day before surgery		
Comparative			
Notes			
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?			
Was a case-control design avoided?			
Did the study avoid inappropriate exclusions?			
DOMAIN 2: Index Test All Tests			

Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Did the assessors of the index test have appropriate training?			
Was interobserver variability reported for some or all patients?			
Was interobserver agreement acceptable?			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?			
Were the reference standard results interpreted without knowledge of the results of the index tests?			
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?			
Did all patients receive the same reference standard?			
Were all patients included in the analysis?			

De Jong 2015

Study characteristics			
Patient sampling	All consecutive intubation procedures in obese (BMI >= 30) patients using two multicentre databases, one containing data from 60 French medical, surgical ICUs, and the other containing data from four anaesthesia departments		
Patient characteristics and setting	Obese (BMI >= 30) patients Exclusion criteria were pregnancy or being under 18 years of age setting both ICU and OR Sample size: 2385 1238 females Mean age: 55 years		
Index tests	MMT, TMD (< 5 cm), mouth opening		
Target condition and reference standard(s)	Difficult tracheal intubation: three or more laryngoscopic attempts or > 10 minutes		
Flow and timing	No information		
Comparative			
Notes			
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?			
Was a case-control design avoided?			
Did the study avoid inappropriate exclusions?			
DOMAIN 2: Index Test All Tests			
Were the index test results interpreted without knowledge of the results of the reference standard?			

De Jong 2015 (*Continued*)

If a threshold was used, was it pre-specified?			
Did the assessors of the index test have appropriate training?			
Was interobserver variability reported for some or all patients?			
Was interobserver agreement acceptable?			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?			
Were the reference standard results interpreted without knowledge of the results of the index tests?			
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?			
Did all patients receive the same reference standard?			
Were all patients included in the analysis?			

Descoins 1994

Study characteristics	
Patient sampling	ENT patients

Descoins 1994 (*Continued*)

Patient characteristics and setting	Sample size: 295		
Index tests	MMT, TMD, mouth opening (< 5 cm), combination of tests		
Target condition and reference standard(s)	Difficult laryngoscopy: Cormack and Lehane		
Flow and timing	No information		
Comparative			
Notes			
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?			
Was a case-control design avoided?			
Did the study avoid inappropriate exclusions?			
DOMAIN 2: Index Test All Tests			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Did the assessors of the index test have appropriate training?			
Was interobserver variability reported for some or all patients?			

Descoins 1994 (*Continued*)

Was interobserver agreement acceptable?			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?			
Were the reference standard results interpreted without knowledge of the results of the index tests?			
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?			
Did all patients receive the same reference standard?			
Were all patients included in the analysis?			

Dohrn 2015

Study characteristics	
Patient sampling	Consecutive patients scheduled for laparoscopic gastric bypass surgery
Patient characteristics and setting	Sample size: 539 437 females
Index tests	MMT, combination of tests
Target condition and reference standard(s)	Difficult tracheal intubation: more than two attempts
Flow and timing	No further information

Dohrn 2015 (*Continued*)

Comparative			
Notes			
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?			
Was a case-control design avoided?			
Did the study avoid inappropriate exclusions?			
DOMAIN 2: Index Test All Tests			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Did the assessors of the index test have appropriate training?			
Was interobserver variability reported for some or all patients?			
Was interobserver agreement acceptable?			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?			

Dohrn 2015 (*Continued*)

Were the reference standard results interpreted without knowledge of the results of the index tests?			
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?			
Did all patients receive the same reference standard?			
Were all patients included in the analysis?			

Domi 2009

Study characteristics			
Patient sampling	"The patients were selected and evaluated by a senior anesthesiologist"		
Patient characteristics and setting	"All the patients with previous anesthesia records sugesting difficult intubation as well as patients with congenital and acquired illnesses of neck and head were excluded from the study" Sample size: 426		
Index tests	MMT, Wilson risk score, TMD, SMD, mouth opening (< 4 cm), combination of tests		
Target condition and reference standard(s)	Difficult laryngoscopy: Cormack and Lehane		
Flow and timing	Unknown		
Comparative			
Notes			
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			

Was a consecutive or random sample of patients enrolled?			
Was a case-control design avoided?			
Did the study avoid inappropriate exclusions?			
DOMAIN 2: Index Test All Tests			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Did the assessors of the index test have appropriate training?			
Was interobserver variability reported for some or all patients?			
Was interobserver agreement acceptable?			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?			
Were the reference standard results interpreted without knowledge of the results of the index tests?			
DOMAIN 4: Flow and Timing			

Domi 2009 (*Continued*)

Was there an appropriate interval between index test and reference standard?			
Did all patients receive the same reference standard?			
Were all patients included in the analysis?			

Domi 2010

Study characteristics			
Patient sampling	Convenience sample		
Patient characteristics and setting	Exclusion: < 14 years; history of difficult to intubate Sample size: 426 209 females		
Index tests	MMT, TMD, SMD		
Target condition and reference standard(s)	Difficult laryngoscopy: Cormack and Lehane		
Flow and timing	Unknown		
Comparative			
Notes			
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?			
Was a case-control design avoided?			
Did the study avoid inappropriate exclusions?			

Domi 2010 (*Continued*)

DOMAIN 2: Index Test All Tests			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Did the assessors of the index test have appropriate training?			
Was interobserver variability reported for some or all patients?			
Was interobserver agreement acceptable?			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?			
Were the reference standard results interpreted without knowledge of the results of the index tests?			
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?			
Did all patients receive the same reference standard?			
Were all patients included in the analysis?			

Eberhart 2005

Study characteristics			
Patient sampling	Consecutive		
Patient characteristics and setting	Sample size: 1269 449 females Mean age: 61 years		
Index tests	MMT, ULBT		
Target condition and reference standard(s)	Difficult laryngoscopy: Cormack and Lehane		
Flow and timing	Unclear		
Comparative			
Notes			
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?			
Was a case-control design avoided?			
Did the study avoid inappropriate exclusions?			
DOMAIN 2: Index Test All Tests			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Did the assessors of the index test have appropriate training?			

Eberhart 2005 (*Continued*)

Was interobserver variability reported for some or all patients?			
Was interobserver agreement acceptable?			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?			
Were the reference standard results interpreted without knowledge of the results of the index tests?			
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?			
Did all patients receive the same reference standard?			
Were all patients included in the analysis?			

El-Ganzouri 1996

Study characteristics	
Patient sampling	All patients who underwent general surgery
Patient characteristics and setting	Sample size: 10,507
Index tests	MMT, TMD (< 6 cm), mouth opening, combination of tests
Target condition and reference standard(s)	Difficult laryngoscopy: Cormack and Lehane

El-Ganzouri 1996 (*Continued*)

Flow and timing	Not specified		
Comparative			
Notes			
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?			
Was a case-control design avoided?			
Did the study avoid inappropriate exclusions?			
DOMAIN 2: Index Test All Tests			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Did the assessors of the index test have appropriate training?			
Was interobserver variability reported for some or all patients?			
Was interobserver agreement acceptable?			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?			

El-Ganzouri 1996 (*Continued*)

Were the reference standard results interpreted without knowledge of the results of the index tests?			
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?			
Did all patients receive the same reference standard?			
Were all patients included in the analysis?			

Ezri 2001

Study characteristics			
Patient sampling	All patients > 18 years in preoperative holding area between 08:00 h to 16:00 h enrolled		
Patient characteristics and setting	Excluded patients given regional anaesthesia and patients receiving GA without endotracheal intubation. Excluded also patients with upper airway pathology, cervical spine fractures and increased risk for aspiration of gastric contents Sample size: 764 367 females Mean age: 44.4 years		
Index tests	MMT		
Target condition and reference standard(s)	Difficult laryngoscopy: Cormack and Lehane		
Flow and timing	Exact time interval between airway bedside test and laryngoscopy not described		
Comparative			
Notes			
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**

Ezri 2001 (*Continued*)

DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?			
Was a case-control design avoided?			
Did the study avoid inappropriate exclusions?			
DOMAIN 2: Index Test All Tests			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Did the assessors of the index test have appropriate training?			
Was interobserver variability reported for some or all patients?			
Was interobserver agreement acceptable?			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?			
Were the reference standard results interpreted without knowledge of the results of the index tests?			
DOMAIN 4: Flow and Timing			

Was there an appropriate interval between index test and reference standard?			
Did all patients receive the same reference standard?			
Were all patients included in the analysis?			

Ezri 2003a

Study characteristics			
Patient sampling	All consecutive patients older than 18 years of age, who arrived in the preoperative holding area for elective surgery		
Patient characteristics and setting	Patients with upper airway pathology, history of difficult laryngoscopy/intubation and full stomach were excluded Sample size: 50 29 females Mean age: 35 years		
Index tests	MMT, TMD (< 6 cm), mouth opening, combination of tests		
Target condition and reference standard(s)	Difficult laryngoscopy: Cormack and Lehane		
Flow and timing	No information		
Comparative			
Notes			
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?			
Was a case-control design avoided?			

Did the study avoid inappropriate exclusions?			
DOMAIN 2: Index Test All Tests			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Did the assessors of the index test have appropriate training?			
Was interobserver variability reported for some or all patients?			
Was interobserver agreement acceptable?			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?			
Were the reference standard results interpreted without knowledge of the results of the index tests?			
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?			
Did all patients receive the same reference standard?			

Ezri 2003a (*Continued*)

Were all patients included in the analysis?			

Ezri 2003b

Study characteristics			
Patient sampling	Consecutive patients undergoing coronary artery bypass surgery and general surgery (laparoscopies and open laparotomies)		
Patient characteristics and setting	All aged > 40 years, patients with BMI > 35, upper airway pathology, history of difficult aryngoscopy/ intubation and full stomach were excluded Sample size: 1472 735 females Mean age: 44.2 years		
Index tests	MMT, TMD (< 6 cm), mouth opening (< 4 cm), combination of tests		
Target condition and reference standard(s)	Difficult laryngoscopy: Cormack and Lehane		
Flow and timing	No information		
Comparative			
Notes			
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?			
Was a case-control design avoided?			
Did the study avoid inappropriate exclusions?			
DOMAIN 2: Index Test All Tests			

Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Did the assessors of the index test have appropriate training?			
Was interobserver variability reported for some or all patients?			
Was interobserver agreement acceptable?			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?			
Were the reference standard results interpreted without knowledge of the results of the index tests?			
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?			
Did all patients receive the same reference standard?			
Were all patients included in the analysis?			

Ezri 2003c

Study characteristics			
Patient sampling	Morbidly obese (BMI > 35) scheduled for weight reduction surgery		
Patient characteristics and setting	See above		
Index tests	MMT		
Target condition and reference standard(s)	Difficult tracheal intubation		
Flow and timing	Night before surgery		
Comparative			
Notes			
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?			
Was a case-control design avoided?			
Did the study avoid inappropriate exclusions?			
DOMAIN 2: Index Test All Tests			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Did the assessors of the index test have appropriate training?			

Ezri 2003c (*Continued*)

Was interobserver variability reported for some or all patients?			
Was interobserver agreement acceptable?			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?			
Were the reference standard results interpreted without knowledge of the results of the index tests?			
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?			
Did all patients receive the same reference standard?			
Were all patients included in the analysis?			

Frerk 1991

Study characteristics	
Patient sampling	Adults requiring tracheal intubation as part of anaesthesia assessed before operation
Patient characteristics and setting	Sample size: 244 101 females Mean age: 44.3 years
Index tests	MMT, TMD (< 7 cm), combination of tests

Frerk 1991 (*Continued*)

Target condition and reference standard(s)	Difficult laryngoscopy: Cormack and Lehane, Macintosh blade for laryngoscopy		
Flow and timing	Tests done at preoperative visit		
Comparative			
Notes			
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?			
Was a case-control design avoided?			
Did the study avoid inappropriate exclusions?			
DOMAIN 2: Index Test All Tests			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Did the assessors of the index test have appropriate training?			
Was interobserver variability reported for some or all patients?			
Was interobserver agreement acceptable?			
DOMAIN 3: Reference Standard			

Frerk 1991 (*Continued*)

Is the reference standards likely to correctly classify the target condition?			
Were the reference standard results interpreted without knowledge of the results of the index tests?			
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?			
Did all patients receive the same reference standard?			
Were all patients included in the analysis?			

Frerk 1996

Study characteristics	
Patient sampling	Case-control
Patient characteristics and setting	Ten patients with a history of difficult tracheal intubation (Cormack and Lehane grade III or IV) and 10 control patients in whom the trachea was easy to intubate (Cormack and Lehane grade I or II) were examined Sample size: 20
Index tests	MMT
Target condition and reference standard(s)	Difficult laryngoscopy: Cormack and Lehane
Flow and timing	Unknown
Comparative	
Notes	
Methodological quality	

Frerk 1996 (*Continued*)

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?			
Was a case-control design avoided?			
Did the study avoid inappropriate exclusions?			
DOMAIN 2: Index Test All Tests			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Did the assessors of the index test have appropriate training?			
Was interobserver variability reported for some or all patients?			
Was interobserver agreement acceptable?			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?			
Were the reference standard results interpreted without knowledge of the results of the index tests?			

Frerk 1996 (*Continued*)

DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?			
Did all patients receive the same reference standard?			
Were all patients included in the analysis?			

Freund 2012

Study characteristics	
Patient sampling	Unknown
Patient characteristics and setting	Patients intubated in physician-staffed EMS; patients with alternative airway management at first or second attempt excluded Sample size: 694 264 females Mean age: 60.5 years
Index tests	TMD, combination of tests
Target condition and reference standard(s)	Difficult laryngoscopy: Cormack and Lehane
Flow and timing	No information
Comparative	
Notes	

Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?			
Was a case-control design avoided?			

Freund 2012 (*Continued*)

Did the study avoid inappropriate exclusions?			
DOMAIN 2: Index Test All Tests			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Did the assessors of the index test have appropriate training?			
Was interobserver variability reported for some or all patients?			
Was interobserver agreement acceptable?			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?			
Were the reference standard results interpreted without knowledge of the results of the index tests?			
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?			
Did all patients receive the same reference standard?			

Were all patients included in the analysis?			

Fritscherova 2011

Study characteristics			
Patient sampling	Patients over 18 years of age undergoing surgery under general anaesthesia with their airway secured by tracheal intubation		
Patient characteristics and setting	Patients in whom difficult intubation could be anticipated were excluded from the study Sample size: 158 78 females Mean age: 59.6 years		
Index tests	MMT, TMD, ULBT, combination of tests		
Target condition and reference standard(s)	Difficult laryngoscopy: Cormack and Lehane; difficult tracheal intubation: failed or > 10 mins		
Flow and timing	Reference standard in the operation Index test the following day for difficult intubation group, no details for easy intubation group		
Comparative			
Notes			
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?			
Was a case-control design avoided?			
Did the study avoid inappropriate exclusions?			
DOMAIN 2: Index Test All Tests			

Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Did the assessors of the index test have appropriate training?			
Was interobserver variability reported for some or all patients?			
Was interobserver agreement acceptable?			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?			
Were the reference standard results interpreted without knowledge of the results of the index tests?			
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?			
Did all patients receive the same reference standard?			
Were all patients included in the analysis?			

Gonzalez 2008

Study characteristics			
Patient sampling	All obese patients scheduled for surgery under general anaesthesia with endotracheal intubation were enrolled in this prospective study at University Hospital of Toulouse Obesity was defined as a BMI > 30 kg/m^2. Concomitantly, all the lean (BMI < 30 kg/m^2) adult patients who were scheduled for surgery during the same period and who were intubated by the same anaesthesiologists were included in the control group		
Patient characteristics and setting	Sample size: 131 115 females		
Index tests	MMT		
Target condition and reference standard(s)	Difficult tracheal intubation: IDS		
Flow and timing	Unknown		
Comparative			
Notes			
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?			
Was a case-control design avoided?			
Did the study avoid inappropriate exclusions?			
DOMAIN 2: Index Test All Tests			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			

Gonzalez 2008 (*Continued*)

Did the assessors of the index test have appropriate training?			
Was interobserver variability reported for some or all patients?			
Was interobserver agreement acceptable?			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?			
Were the reference standard results interpreted without knowledge of the results of the index tests?			
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?			
Did all patients receive the same reference standard?			
Were all patients included in the analysis?			

Hagberg 2009

Study characteristics	
Patient sampling	Retrospective analysis
Patient characteristics and setting	Obese (BMI > 35) patients undergoing elective surgery during a period of 9 years within one hospital Sample size: 283 216 females Mean age: 44.6 years

Hagberg 2009 (*Continued*)

Index tests	MMT		
Target condition and reference standard(s)	Failed intubation		
Flow and timing	Unclear		
Comparative			
Notes			
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?			
Was a case-control design avoided?			
Did the study avoid inappropriate exclusions?			
DOMAIN 2: Index Test All Tests			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Did the assessors of the index test have appropriate training?			
Was interobserver variability reported for some or all patients?			
Was interobserver agreement acceptable?			

Hagberg 2009 (*Continued*)

DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?			
Were the reference standard results interpreted without knowledge of the results of the index tests?			
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?			
Did all patients receive the same reference standard?			
Were all patients included in the analysis?			

Hagiwara 2015

Study characteristics	
Patient sampling	Patients requiring emergency intubation at the ED Those where alternative airway techniques at first attempt were excluded
Patient characteristics and setting	Sample size: 3313 1236 females Mean age: 71 years
Index tests	Combination of tests
Target condition and reference standard(s)	Difficult tracheal intubation: more than two attempts
Flow and timing	Unclear; form filled out after intubation
Comparative	
Notes	

Hagiwara 2015 (*Continued*)

Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?			
Was a case-control design avoided?			
Did the study avoid inappropriate exclusions?			
DOMAIN 2: Index Test All Tests			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Did the assessors of the index test have appropriate training?			
Was interobserver variability reported for some or all patients?			
Was interobserver agreement acceptable?			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?			
Were the reference standard results interpreted without knowledge of the results of the index tests?			

Hagiwara 2015 (*Continued*)

DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?			
Did all patients receive the same reference standard?			
Were all patients included in the analysis?			

Hashim 2014

Study characteristics			
Patient sampling	Patients who were diabetic for at least a year in the age group between 30 and 80 years and underwent elective surgery under general anaesthesia with endotracheal intubation		
Patient characteristics and setting	Patients with obvious anatomical variation of their face, neck, palate or hands and history of difficult intubation in the past were excluded from the study. Patients with coexisting diseases such as rheumatoid arthritis, oral malignancies and large neck masses were also excluded Sample size: 60 37 females Mean age: 56 years		
Index tests	MMT, TMD (< 6 cm), combination of tests		
Target condition and reference standard(s)	Difficult laryngoscopy: Cormack and Lehane using Macintosh blade		
Flow and timing	Preoperatively		
Comparative			
Notes			
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			

Was a consecutive or random sample of patients enrolled?			
Was a case-control design avoided?			
Did the study avoid inappropriate exclusions?			
DOMAIN 2: Index Test All Tests			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Did the assessors of the index test have appropriate training?			
Was interobserver variability reported for some or all patients?			
Was interobserver agreement acceptable?			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?			
Were the reference standard results interpreted without knowledge of the results of the index tests?			
DOMAIN 4: Flow and Timing			

Hashim 2014 (*Continued*)

Was there an appropriate interval between index test and reference standard?			
Did all patients receive the same reference standard?			
Were all patients included in the analysis?			

Healy 2016

Study characteristics			
Patient sampling	Patients undergoing general anaesthesia with a documented preoperative airway examination in combination with a documented glottic view obtained at direct laryngoscopy		
Patient characteristics and setting	Excluded all patients without a documented intraoperative view or presence of an existing airway and patients in which intubation was performed by alternative means Sample size: 80,709 43,015 females		
Index tests	MMT, combination of tests		
Target condition and reference standard(s)	Difficult laryngoscopy: Cormack and Lehane, using either Macintosh or Miller blade Difficult tracheal intubation: IDS		
Flow and timing	Unclear		
Comparative			
Notes			
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?			
Was a case-control design avoided?			

Did the study avoid inappropriate exclusions?			
DOMAIN 2: Index Test All Tests			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Did the assessors of the index test have appropriate training?			
Was interobserver variability reported for some or all patients?			
Was interobserver agreement acceptable?			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?			
Were the reference standard results interpreted without knowledge of the results of the index tests?			
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?			
Did all patients receive the same reference standard?			

Healy 2016 (*Continued*)

Were all patients included in the analysis?			

Heinrich 2013

Study characteristics			
Patient sampling	Database		
Patient characteristics and setting	Patients undergoing anaesthesia Patients with videolaryngoscopic assistance without documentation of a direct laryngeal view were excluded Sample size: 102,305 50,608 females Mean age: 57 years		
Index tests	MMT		
Target condition and reference standard(s)	Difficult laryngoscopy: Cormack and Lehane, standard cold light MacIntosh blade sized appropriately		
Flow and timing	Preoperative maximum 12 hours		
Comparative			
Notes			
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?			
Was a case-control design avoided?			
Did the study avoid inappropriate exclusions?			
DOMAIN 2: Index Test All Tests			

Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Did the assessors of the index test have appropriate training?			
Was interobserver variability reported for some or all patients?			
Was interobserver agreement acceptable?			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?			
Were the reference standard results interpreted without knowledge of the results of the index tests?			
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?			
Did all patients receive the same reference standard?			
Were all patients included in the analysis?			

Hekiert 2007

Study characteristics			
Patient sampling	Retrospetive analysis of consecutive obese patients		
Patient characteristics and setting	Obese patients only (BMI > 30) Sample size: 14 9 females Mean age: 52.2 years		
Index tests	MMT		
Target condition and reference standard(s)	Difficult laryngoscopy: Cormack and Lehane		
Flow and timing	Reference standard: otolaryngology office index test: OP		
Comparative			
Notes			
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?			
Was a case-control design avoided?			
Did the study avoid inappropriate exclusions?			
DOMAIN 2: Index Test All Tests			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Did the assessors of the index test have appropriate training?			

Hekiert 2007 (*Continued*)

Was interobserver variability reported for some or all patients?			
Was interobserver agreement acceptable?			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?			
Were the reference standard results interpreted without knowledge of the results of the index tests?			
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?			
Did all patients receive the same reference standard?			
Were all patients included in the analysis?			

Hirmanpour 2014

Study characteristics	
Patient sampling	Unclear
Patient characteristics and setting	Patients with a history of trauma to the airway or cranial, cervical and facial regions, or were edentulous or requiring awake intubation, patients with restricted motility of the neck and mandible (e.g. cervical disc disorders or rheumatoid arthritis) and inability to sit were not included in the study Sample size: 657 657 females Mean age: 27 years

Hirmanpour 2014 (*Continued*)

Index tests	MMT, ULBT		
Target condition and reference standard(s)	Difficult laryngoscopy: Cormack and Lehane, size three Macintosh laryngoscope blade		
Flow and timing	Preoperative		
Comparative			
Notes			
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?			
Was a case-control design avoided?			
Did the study avoid inappropriate exclusions?			
DOMAIN 2: Index Test All Tests			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Did the assessors of the index test have appropriate training?			
Was interobserver variability reported for some or all patients?			
Was interobserver agreement acceptable?			

Hirmanpour 2014 (*Continued*)

DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?			
Were the reference standard results interpreted without knowledge of the results of the index tests?			
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?			
Did all patients receive the same reference standard?			
Were all patients included in the analysis?			

Honarmand 2008

Study characteristics	
Patient sampling	Consecutive patients selected for elective caesarean delivery
Patient characteristics and setting	Exclusion: < 18, obvious malformations of the airway, inability to sit, head/neck surgery Sample size: 400 400 females Mean age: 24 years
Index tests	MMT, ULBT, combination of tests
Target condition and reference standard(s)	Difficult laryngoscopy: Cormack and Lehane. A anesthesiologist with 7 years experience in anaesthesia, who was not informed of the results of the index tests, carried out laryngoscopy and assessed difficulty of laryngoscopy at intubation, which was performed with the patient adequately anaesthetized and fully relaxed on the operating room table. Laryngoscopy was performed using a Macintosh #4
Flow and timing	Test was carried out prior to transfer to operating room

Honarmand 2008 (*Continued*)

Comparative			
Notes			
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?			
Was a case-control design avoided?			
Did the study avoid inappropriate exclusions?			
DOMAIN 2: Index Test All Tests			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Did the assessors of the index test have appropriate training?			
Was interobserver variability reported for some or all patients?			
Was interobserver agreement acceptable?			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?			

Honarmand 2008 (*Continued*)

Were the reference standard results interpreted without knowledge of the results of the index tests?			
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?			
Did all patients receive the same reference standard?			
Were all patients included in the analysis?			

Honarmand 2014

Study characteristics	
Patient sampling	ASA I-III adult patients programmed to be given general anaesthesia necessitate endotracheal intubation for elective surgery
Patient characteristics and setting	Patients with a history of previous surgery, burns or trauma to the airways or to the cranial, cervical and facial regions, patients with tumours or a mass in the above-mentioned regions, patients with restricted motility of the neck and mandible (e.g. rheumatoid arthritis or cervical disk disorders), inability to sit, edentulous or need awake intubation were excluded from the study Sample size: 525 184 females Mean age: 46 years
Index tests	MMT, ULBT
Target condition and reference standard(s)	Difficult laryngoscopy: Cormack and Lehane, laryngoscopy was done with using a Macintosh #4 blade to visualize the larynx
Flow and timing	Preoperative
Comparative	
Notes	
Methodological quality	

Honarmand 2014 (*Continued*)

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?			
Was a case-control design avoided?			
Did the study avoid inappropriate exclusions?			
DOMAIN 2: Index Test All Tests			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Did the assessors of the index test have appropriate training?			
Was interobserver variability reported for some or all patients?			
Was interobserver agreement acceptable?			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?			
Were the reference standard results interpreted without knowledge of the results of the index tests?			

Honarmand 2014 (*Continued*)

DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?			
Did all patients receive the same reference standard?			
Were all patients included in the analysis?			

Honarmand 2015

Study characteristics	
Patient sampling	Adult patients, who were scheduled to undergo elective operations under general anaesthesia with endotracheal intubation
Patient characteristics and setting	Sample size: 600 319 females Mean age: 44 years
Index tests	MMT, ULBT
Target condition and reference standard(s)	Difficult laryngoscopy: Cormack and Lehane. Laryngoscopy was done by a Macintosh number 4 laryngoscope blade
Flow and timing	Preoperative
Comparative	
Notes	

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?			
Was a case-control design avoided?			

Did the study avoid inappropriate exclusions?			
DOMAIN 2: Index Test All Tests			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Did the assessors of the index test have appropriate training?			
Was interobserver variability reported for some or all patients?			
Was interobserver agreement acceptable?			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?			
Were the reference standard results interpreted without knowledge of the results of the index tests?			
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?			
Did all patients receive the same reference standard?			

Were all patients included in the analysis?			

Huh 2009

Study characteristics			
Patient sampling	Consecutive adult patients scheduled to undergo general anesthesia requiring tracheal intubation for elective surgery		
Patient characteristics and setting	Exclusion criteria included a gross anatomical abnormality, recent surgery of the head and neck, upper airway disease (e.g. maxillofacial fracture or tumours), loose teeth, or those requiring a rapid sequence or awake intubation Sample size: 213 104 females		
Index tests	MMT, TMD (< 6.2 cm)		
Target condition and reference standard(s)	Difficult laryngoscopy: Cormack and Lehane		
Flow and timing	Reference standard immediately after index tests		
Comparative			
Notes			
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?			
Was a case-control design avoided?			
Did the study avoid inappropriate exclusions?			
DOMAIN 2: Index Test All Tests			

Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Did the assessors of the index test have appropriate training?			
Was interobserver variability reported for some or all patients?			
Was interobserver agreement acceptable?			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?			
Were the reference standard results interpreted without knowledge of the results of the index tests?			
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?			
Did all patients receive the same reference standard?			
Were all patients included in the analysis?			

Hui 2009

Study characteristics			
Patient sampling	Patients presenting for elective, non-cardiac surgery requiring intubation		
Patient characteristics and setting	Sample size: 27		
Index tests	MT		
Target condition and reference standard(s)	Difficult laryngoscopy: Cormack and Lehane		
Flow and timing	No information		
Comparative			
Notes			
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?			
Was a case-control design avoided?			
Did the study avoid inappropriate exclusions?			
DOMAIN 2: Index Test All Tests			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Did the assessors of the index test have appropriate training?			

Was interobserver variability reported for some or all patients?			
Was interobserver agreement acceptable?			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?			
Were the reference standard results interpreted without knowledge of the results of the index tests?			
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?			
Did all patients receive the same reference standard?			
Were all patients included in the analysis?			

Istvan 2010

Study characteristics	
Patient sampling	Retrospective chart review of patients undergoing appendectomy within 1 year
Patient characteristics and setting	Inclusion criteria were patients from all ages and sexes who were admitted to hospital from the emergency department and whose preoperative and postoperative diagnosis was acute appendicitis. Exclusion criteria were patients already in hospital whose postoperative diagnosis was not acute appendicitis or who underwent other surgical procedures in the same setting Sample size: 254 100 females Mean age: 29.5 years

Index tests	MMT		
Target condition and reference standard(s)	Failed intubation		
Flow and timing	Unclear		
Comparative			
Notes			
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?			
Was a case-control design avoided?			
Did the study avoid inappropriate exclusions?			
DOMAIN 2: Index Test All Tests			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Did the assessors of the index test have appropriate training?			
Was interobserver variability reported for some or all patients?			
Was interobserver agreement acceptable?			

Istvan 2010 (*Continued*)

DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?			
Were the reference standard results interpreted without knowledge of the results of the index tests?			
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?			
Did all patients receive the same reference standard?			
Were all patients included in the analysis?			

Ittichaikulthol 2010

Study characteristics	
Patient sampling	Consecutive ASA I-IV adult patients
Patient characteristics and setting	Patients scheduled to receive general anaesthesia requiring endotracheal intubation for all surgery Sample size: 1888 1239 females
Index tests	MMT (I versus II-IV), TMD (< 6 cm), combination of tests
Target condition and reference standard(s)	Difficult laryngoscopy: Cormack and Lehane
Flow and timing	Unclear
Comparative	
Notes	

Ittichaikulthol 2010 (*Continued*)

Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?			
Was a case-control design avoided?			
Did the study avoid inappropriate exclusions?			
DOMAIN 2: Index Test All Tests			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Did the assessors of the index test have appropriate training?			
Was interobserver variability reported for some or all patients?			
Was interobserver agreement acceptable?			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?			
Were the reference standard results interpreted without knowledge of the results of the index tests?			

Ittichaikulthol 2010 (*Continued*)

DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?			
Did all patients receive the same reference standard?			
Were all patients included in the analysis?			

Juvin 2003

Study characteristics			
Patient sampling	All obese (BMI > 35), adult (> 18 years) patients scheduled for laparoscopic gastroplasty and all lean (BMI < 30) adult patients who were scheduled for inguinal hernia repair or laparoscopic cholecystectomy during the same period and who were intubated by the same anaesthesiologists were included		
Patient characteristics and setting	Excluded: ASA III or IV, BMI 30 to 35 Sample size: 263 189 females Mean age: 41 years		
Index tests	MMT, mouth opening, combination of tests		
Target condition and reference standard(s)	Difficult tracheal intubation: IDS		
Flow and timing	No information		
Comparative			
Notes			
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			

Was a consecutive or random sample of patients enrolled?			
Was a case-control design avoided?			
Did the study avoid inappropriate exclusions?			
DOMAIN 2: Index Test All Tests			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Did the assessors of the index test have appropriate training?			
Was interobserver variability reported for some or all patients?			
Was interobserver agreement acceptable?			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?			
Were the reference standard results interpreted without knowledge of the results of the index tests?			
DOMAIN 4: Flow and Timing			

Juvin 2003 (*Continued*)

Was there an appropriate interval between index test and reference standard?			
Did all patients receive the same reference standard?			
Were all patients included in the analysis?			

K Nasa 2014

Study characteristics			
Patient sampling	Patients above the age of 12 years who were fit for general endotracheal anaesthesia irrespective of their ASA physical status were included in the study		
Patient characteristics and setting	Patients with obvious airway malformations, patient with inter incisor distance < 3 cm, patients allergic to drugs used in study were excluded from the study Sample size: 400 190 females		
Index tests	MMT, TMD (< 6 cm), combination of tests		
Target condition and reference standard(s)	Difficult tracheal intubation: IDS		
Flow and timing	Unknown		
Comparative			
Notes			
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?			
Was a case-control design avoided?			

Did the study avoid inappropriate exclusions?			
DOMAIN 2: Index Test All Tests			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Did the assessors of the index test have appropriate training?			
Was interobserver variability reported for some or all patients?			
Was interobserver agreement acceptable?			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?			
Were the reference standard results interpreted without knowledge of the results of the index tests?			
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?			
Did all patients receive the same reference standard?			

Were all patients included in the analysis?			

Kalezic 2016

Study characteristics			
Patient sampling	Consecutive adult patients scheduled for thyroid surgery		
Patient characteristics and setting	Sample size: 2000 1705 females		
Index tests	MMT, combination of tests		
Target condition and reference standard(s)	Difficult laryngoscopy: Cormack and Lehane; difficult tracheal intubation		
Flow and timing	Unknown		
Comparative			
Notes			
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?			
Was a case-control design avoided?			
Did the study avoid inappropriate exclusions?			
DOMAIN 2: Index Test All Tests			
Were the index test results interpreted without knowledge of the results of the reference standard?			

Kalezic 2016 (*Continued*)

If a threshold was used, was it pre-specified?			
Did the assessors of the index test have appropriate training?			
Was interobserver variability reported for some or all patients?			
Was interobserver agreement acceptable?			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?			
Were the reference standard results interpreted without knowledge of the results of the index tests?			
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?			
Did all patients receive the same reference standard?			
Were all patients included in the analysis?			

Kamalipour 2005

Study characteristics	
Patient sampling	Adult patients who were scheduled for elective surgery under general anaesthesia were randomly selected (using the branched block randomization method) and considered for enrolment

Kamalipour 2005 (*Continued*)

Patient characteristics and setting	Patients with obvious malformations of the airway, edentulous patients, patients who required cricoid pressure for rapid sequence intubation and pregnant women were excluded from the study. Edentulous patients were also excluded Sample size: 100 36 females		
Index tests	MMT		
Target condition and reference standard(s)	Difficult laryngoscopy: Cormack and Lehane		
Flow and timing	No information		
Comparative			
Notes			

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?			
Was a case-control design avoided?			
Did the study avoid inappropriate exclusions?			
DOMAIN 2: Index Test All Tests			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Did the assessors of the index test have appropriate training?			
Was interobserver variability reported for some or all patients?			

Kamalipour 2005 (*Continued*)

Was interobserver agreement acceptable?			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?			
Were the reference standard results interpreted without knowledge of the results of the index tests?			
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?			
Did all patients receive the same reference standard?			
Were all patients included in the analysis?			

Kamranmanesh 2013

Study characteristics	
Patient sampling	Consecutive adult asian patients aged 20 to 65 years with ASA I and II, scheduled to undergo elective surgery requiring endotracheal intubation, were enrolled in this prospective observational study
Patient characteristics and setting	Exclusion criteria were as follows: obvious anatomical abnormality, upper airway abnormality (e.g. tong tumour, maxillofacial tumour, or fracture), recent head and neck surgery, ASA class III and IV, and disability to open the mouth Sample size: 603 173 females Mean age: 42.4 years
Index tests	MMT, combination of tests

Target condition and reference standard(s)	Difficult laryngoscopy: Cormack and Lehane		
Flow and timing	Unclear		
Comparative			
Notes			
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?			
Was a case-control design avoided?			
Did the study avoid inappropriate exclusions?			
DOMAIN 2: Index Test All Tests			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Did the assessors of the index test have appropriate training?			
Was interobserver variability reported for some or all patients?			
Was interobserver agreement acceptable?			
DOMAIN 3: Reference Standard			

Is the reference standards likely to correctly classify the target condition?			
Were the reference standard results interpreted without knowledge of the results of the index tests?			
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?			
Did all patients receive the same reference standard?			
Were all patients included in the analysis?			

Khan 2003

Study characteristics	
Patient sampling	Consecutive male and female patients, aged >= 16 years, scheduled to undergo surgery under general anaesthesia between January 2001 and November 2001, were considered for enrolment
Patient characteristics and setting	Edentulous patients, those unable to open the mouth, with laryngeal masses, or with limitation of cervical movements were excluded from the study Sample size: 300
Index tests	MMT, ULBT
Target condition and reference standard(s)	Difficult laryngoscopy: Cormack and Lehane
Flow and timing	No information
Comparative	
Notes	
Methodological quality	

Khan 2003 (*Continued*)

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?			
Was a case-control design avoided?			
Did the study avoid inappropriate exclusions?			
DOMAIN 2: Index Test All Tests			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Did the assessors of the index test have appropriate training?			
Was interobserver variability reported for some or all patients?			
Was interobserver agreement acceptable?			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?			
Were the reference standard results interpreted without knowledge of the results of the index tests?			

Khan 2003 (*Continued*)

DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?			
Did all patients receive the same reference standard?			
Were all patients included in the analysis?			

Khan 2009a

Study characteristics			
Patient sampling	ASA I patients older than 16 years scheduled for elective surgical procedures requiring endotracheal intubation were enrolled		
Patient characteristics and setting	Patients with any airway abnormality or obvious neck pathology were excluded Sample size: 380 171 females Mean age: 34 years		
Index tests	TMD (< 13.5 cm), SMD, mouth opening, ULBT, combination of tests		
Target condition and reference standard(s)	Difficult laryngoscopy: Cormack and Lehane		
Flow and timing	Unknown		
Comparative			
Notes			
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?			
Was a case-control design avoided?			

Did the study avoid inappropriate exclusions?			
DOMAIN 2: Index Test All Tests			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Did the assessors of the index test have appropriate training?			
Was interobserver variability reported for some or all patients?			
Was interobserver agreement acceptable?			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?			
Were the reference standard results interpreted without knowledge of the results of the index tests?			
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?			
Did all patients receive the same reference standard?			

Were all patients included in the analysis?			

Khan 2009b

Study characteristics			
Patient sampling	Patients undergoing surgery and requiring endotracheal intubation were enrolled in this study		
Patient characteristics and setting	Exclusion criteria included compromised critical airway, emergent cases, noncompliable patients and those with anatomical anomalies in the airway, pregnant, edentulous, those having beard and patients less than 14 years and those in whom a good mask fit was not possible Sample size: 200 118 females		
Index tests	MMT, ULBT		
Target condition and reference standard(s)	Difficult face mask ventilation: mask ventilation was performed by means of an appropriate sized face mask applied to the face and a reservoir bag receiving a continuous flow of oxygen from the anaesthesia machine		
Flow and timing	Shortly one after another		
Comparative			
Notes			
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?			
Was a case-control design avoided?			
Did the study avoid inappropriate exclusions?			
DOMAIN 2: Index Test All Tests			

Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Did the assessors of the index test have appropriate training?			
Was interobserver variability reported for some or all patients?			
Was interobserver agreement acceptable?			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?			
Were the reference standard results interpreted without knowledge of the results of the index tests?			
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?			
Did all patients receive the same reference standard?			
Were all patients included in the analysis?			

Khan 2011

Study characteristics			
Patient sampling	Consecutive patients of ASA physical status I and II, aged 20-60 scheduled for elective surgical procedures requiring tracheal intubation between July 2008 and June 2009		
Patient characteristics and setting	Edentulous patients, those unable to open the mouth, patients with pharyngolaryngeal pathology, with a history of thyroid neck surgery, pregnancy, or with limitation of temporomandibular and atlanto-axial joints were excluded from the study Sample size: 300 175 females Mean age: 38.4 years		
Index tests	ULBT, combination of tests		
Target condition and reference standard(s)	Difficult laryngoscopy: Cormack and Lehane		
Flow and timing	Preoperatively		
Comparative			
Notes			
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?			
Was a case-control design avoided?			
Did the study avoid inappropriate exclusions?			
DOMAIN 2: Index Test All Tests			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			

Khan 2011 (*Continued*)

Did the assessors of the index test have appropriate training?			
Was interobserver variability reported for some or all patients?			
Was interobserver agreement acceptable?			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?			
Were the reference standard results interpreted without knowledge of the results of the index tests?			
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?			
Did all patients receive the same reference standard?			
Were all patients included in the analysis?			

Khan 2013

Study characteristics	
Patient sampling	Consecutive patients, ASA I to III who required general anaesthesia and endotracheal intubation were studied prospectively over a 3- year period from January 2007 until December 2010
Patient characteristics and setting	Exclusion criteria included inability to sit, gross anatomical abnormality or recent surgery of the head and neck and patients with pregnancy or severe cardiorespiratory disorders Sample size: 4500

Khan 2013 (*Continued*)

	1505 females Mean age: 55.7 years		
Index tests	TMD, SMD, ULBT		
Target condition and reference standard(s)	Difficult laryngoscopy: Cormack and Lehane		
Flow and timing	Unclear		
Comparative			
Notes			
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?			
Was a case-control design avoided?			
Did the study avoid inappropriate exclusions?			
DOMAIN 2: Index Test All Tests			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Did the assessors of the index test have appropriate training?			
Was interobserver variability reported for some or all patients?			
Was interobserver agreement acceptable?			

Khan 2013 (*Continued*)

DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?			
Were the reference standard results interpreted without knowledge of the results of the index tests?			
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?			
Did all patients receive the same reference standard?			
Were all patients included in the analysis?			

Khan 2014

Study characteristics	
Patient sampling	Consecutive male or female edentulous patients ≥ 60 years old scheduled to undergo elective surgery under GA between March 2008 and June 2011 were considered for enrolment
Patient characteristics and setting	Uncooperative patients, those unable to open the mouth or with pharyngolaryngeal pathology were excluded from the study. Patients with fixed prosthetic dentures were also excluded and mobile dentures, if present, were removed to adhere to the true definition of an edentulous case Sample size: 588 253 females Mean age: 64 years
Index tests	MMT, combination of tests
Target condition and reference standard(s)	Difficult laryngoscopy: Cormack and Lehane

Flow and timing	Unclear		
Comparative			
Notes			
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?			
Was a case-control design avoided?			
Did the study avoid inappropriate exclusions?			
DOMAIN 2: Index Test All Tests			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Did the assessors of the index test have appropriate training?			
Was interobserver variability reported for some or all patients?			
Was interobserver agreement acceptable?			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?			

Khan 2014 (*Continued*)

Were the reference standard results interpreted without knowledge of the results of the index tests?			
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?			
Did all patients receive the same reference standard?			
Were all patients included in the analysis?			

Khan 2015

Study characteristics			
Patient sampling	"In this cross-sectional study, 661 patients aged 16-60 years were recruited during the years 2011 to 2012"		
Patient characteristics and setting	Exclusion criteria included ASA class higher than II, urgency of the situation, facial, mouth, throat and airway anomalies, pregnancy and awake intubation Sample size: 661 366 females		
Index tests	MMT		
Target condition and reference standard(s)	Difficult laryngoscopy: Cormack and Lehane; difficult tracheal intubation		
Flow and timing	Unclear		
Comparative			
Notes			
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**

Khan 2015 (*Continued*)

DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?			
Was a case-control design avoided?			
Did the study avoid inappropriate exclusions?			
DOMAIN 2: Index Test All Tests			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Did the assessors of the index test have appropriate training?			
Was interobserver variability reported for some or all patients?			
Was interobserver agreement acceptable?			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?			
Were the reference standard results interpreted without knowledge of the results of the index tests?			
DOMAIN 4: Flow and Timing			

Was there an appropriate interval between index test and reference standard?			
Did all patients receive the same reference standard?			
Were all patients included in the analysis?			

Kheterpal 2009

Study characteristics			
Patient sampling	All adult patients (age 18 years or older) undergoing a general anaesthetic at a tertiary care university hospital were included over a 4-year period from 2004 to 2008		
Patient characteristics and setting	All cases without an attempt at mask ventilation were excluded from the data collection and analysis, including planned awake fiberoptic intubations Sample size: 53,041 28,657 females Mean age: 51 years		
Index tests	MMT, TMD (< 6 cm), mouth opening (< 3 cm)		
Target condition and reference standard(s)	Difficult face mask ventilation: inability to establish face mask ventilation despite multiple airway adjuvants and two-hand mask ventilation		
Flow and timing	Unclear		
Comparative			
Notes			
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?			
Was a case-control design avoided?			

Did the study avoid inappropriate exclusions?			
DOMAIN 2: Index Test All Tests			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Did the assessors of the index test have appropriate training?			
Was interobserver variability reported for some or all patients?			
Was interobserver agreement acceptable?			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?			
Were the reference standard results interpreted without knowledge of the results of the index tests?			
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?			
Did all patients receive the same reference standard?			

Kheterpal 2009 (*Continued*)

Were all patients included in the analysis?			

Kim 2011

Study characteristics			
Patient sampling	Patients undergoing surgery under general anaesthesia with tracheal intubation		
Patient characteristics and setting	Patients were divided into obese (BMI >= 27.5) and non-obese groups. Sufficient measures of DTA presented for obese patients only. Same number of obese and non-obese patients (130 each), so consecutive sample is unlikely Sample size: 123 77 females Mean age: 48.6 years		
Index tests	MMT, Wilson risk score, combination of tests		
Target condition and reference standard(s)	Difficult tracheal intubation: IDS		
Flow and timing	No details given		
Comparative			
Notes			
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?			
Was a case-control design avoided?			
Did the study avoid inappropriate exclusions?			
DOMAIN 2: Index Test All Tests			

Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Did the assessors of the index test have appropriate training?			
Was interobserver variability reported for some or all patients?			
Was interobserver agreement acceptable?			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?			
Were the reference standard results interpreted without knowledge of the results of the index tests?			
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?			
Did all patients receive the same reference standard?			
Were all patients included in the analysis?			

Knudsen 2014

Study characteristics			
Patient sampling	Patients scheduled for elective day surgery, inclusion criteria were patients with ASA scores of I or II who were older than 17 years and were scheduled for general anaesthesia requiring endotracheal intubation		
Patient characteristics and setting	Exclusion criteria were anaesthesia with rapid sequence induction, pregnancy, and BMI higher than 35 kg/ma Sample size: 87 68 females Mean age: 42 years		
Index tests	MMT, TMD (< 10 cm), combination of tests		
Target condition and reference standard(s)	Difficult tracheal intubation: "according to ASA recommendations"		
Flow and timing	Preoperatively		
Comparative			
Notes			
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?			
Was a case-control design avoided?			
Did the study avoid inappropriate exclusions?			
DOMAIN 2: Index Test All Tests			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			

Knudsen 2014 (*Continued*)

Did the assessors of the index test have appropriate training?			
Was interobserver variability reported for some or all patients?			
Was interobserver agreement acceptable?			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?			
Were the reference standard results interpreted without knowledge of the results of the index tests?			
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?			
Did all patients receive the same reference standard?			
Were all patients included in the analysis?			

Koh 2002

Study characteristics	
Patient sampling	Succesive adult (> 16 years) patients scheduled for elective surgery under general anaesthesia Exclusion: RSI
Patient characteristics and setting	Sample size: 605 339 females Mean age: 44.5 years

Koh 2002 (*Continued*)

Index tests	MMT, TMD (< 6 cm), combination of tests		
Target condition and reference standard(s)	Difficult laryngoscopy: Cormack and Lehane, Macintosh size 3, best view, BURP if needed; difficult tracheal intubation		
Flow and timing	Unclear		
Comparative			
Notes			
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?			
Was a case-control design avoided?			
Did the study avoid inappropriate exclusions?			
DOMAIN 2: Index Test All Tests			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Did the assessors of the index test have appropriate training?			
Was interobserver variability reported for some or all patients?			
Was interobserver agreement acceptable?			

Koh 2002 (*Continued*)

DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?			
Were the reference standard results interpreted without knowledge of the results of the index tests?			
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?			
Did all patients receive the same reference standard?			
Were all patients included in the analysis?			

Kolarkar 2015

Study characteristics	
Patient sampling	300 patients of either sex, undergoing elective surgery under general anaesthesia with endotracheal intubation. Inclusion criteria being patients of ASA grade I/II, age: 21-60 years of either sex, elective surgery under GA
Patient characteristics and setting	Exclusion criteria were edentulous patient, unable to open the mouth, with pharyngolaryngeal pathology, history of thyroid/neck surgery, limitations of temporomandibular or atlanto-axial joint. Congenital facial deformity Sample size: 300 160 females Mean age: 40.6 years
Index tests	ULBT, combination of tests
Target condition and reference standard(s)	Difficult laryngoscopy: Cormack and Lehane
Flow and timing	Unclear

Comparative			
Notes			
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?			
Was a case-control design avoided?			
Did the study avoid inappropriate exclusions?			
DOMAIN 2: Index Test All Tests			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Did the assessors of the index test have appropriate training?			
Was interobserver variability reported for some or all patients?			
Was interobserver agreement acceptable?			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?			

Kolarkar 2015 (*Continued*)

Were the reference standard results interpreted without knowledge of the results of the index tests?			
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?			
Did all patients receive the same reference standard?			
Were all patients included in the analysis?			

Komatsu 2007

Study characteristics			
Patient sampling	Morbidly obese patients (BMI > 35) scheduled for elective surgery under GA with tracheal intubation. Patients with removable upper dentures, upper airway pathology, cervical spine fractures, full stomach, significant gastro-oesophageal reflux or a history of difficult laryngoscopy were excluded. Pregnant women were also excluded		
Patient characteristics and setting	Sample size: 64 64 females		
Index tests	MMT (I versus II-IV)		
Target condition and reference standard(s)	Difficult laryngoscopy: Cormack and Lehane		
Flow and timing	Unclear		
Comparative			
Notes			
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**

Komatsu 2007 (*Continued*)

DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?			
Was a case-control design avoided?			
Did the study avoid inappropriate exclusions?			
DOMAIN 2: Index Test All Tests			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Did the assessors of the index test have appropriate training?			
Was interobserver variability reported for some or all patients?			
Was interobserver agreement acceptable?			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?			
Were the reference standard results interpreted without knowledge of the results of the index tests?			
DOMAIN 4: Flow and Timing			

Was there an appropriate interval between index test and reference standard?			
Did all patients receive the same reference standard?			
Were all patients included in the analysis?			

Konwar 2015

Study characteristics			
Patient sampling	200 patients were randomly selected and enrolled in this study. The study population consisted of patients of ASA class I and II, belonging to either sex of age group 18-40 years admitted for operation under GA with endotracheal intubation		
Patient characteristics and setting	Patients with concurrent pregnancy; intraoral, laryngeal or pharyngeal mass; altered head and neck anatomy; and restricted movement of the neck were excluded Sample size: 200 83 females Mean age: 28 years		
Index tests	TMD, mouth opening, ULBT		
Target condition and reference standard(s)	Difficult laryngoscopy: Cormack and Lehane		
Flow and timing	Preoperatively		
Comparative			
Notes			
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?			
Was a case-control design avoided?			

Did the study avoid inappropriate exclusions?			
DOMAIN 2: Index Test All Tests			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Did the assessors of the index test have appropriate training?			
Was interobserver variability reported for some or all patients?			
Was interobserver agreement acceptable?			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?			
Were the reference standard results interpreted without knowledge of the results of the index tests?			
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?			
Did all patients receive the same reference standard?			

Were all patients included in the analysis?			

Krobbuaban 2005

Study characteristics			
Patient sampling	Consecutive ASA I-II adult patients scheduled to receive GA requiring endotracheal intubation for elective orthopaedic, urologic, abdominal, and gynaecologic surgery		
Patient characteristics and setting	Patients younger than 18 years of age, with obvious malformations of the airway, edentulous, or requiring a RSI or awake intubation were excluded from the study Sample size: 550 289 females Mean age: 45 years		
Index tests	MMT, TMD, mouth opening		
Target condition and reference standard(s)	Difficult laryngoscopy: Cormack and Lehane		
Flow and timing	Unknown		
Comparative			
Notes			
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?			
Was a case-control design avoided?			
Did the study avoid inappropriate exclusions?			
DOMAIN 2: Index Test All Tests			

Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Did the assessors of the index test have appropriate training?			
Was interobserver variability reported for some or all patients?			
Was interobserver agreement acceptable?			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?			
Were the reference standard results interpreted without knowledge of the results of the index tests?			
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?			
Did all patients receive the same reference standard?			
Were all patients included in the analysis?			

Langeron 2000

Study characteristics			
Patient sampling	All adult patients scheduled for orthopaedic, urologic, abdominal, gynaecologic and neurosurgery with GA were prospectively included in the study over a 6-month period		
Patient characteristics and setting	Those with contraindication of mask ventilation (i.e. emergency cases requiring a RSI, planned awake intubation) were excluded Sample size: 1502 831 females Mean age: 50.5 years		
Index tests	MMT, combination of tests		
Target condition and reference standard(s)	Difficult face mask ventilation: the inability of an unassisted anaesthesiologist to maintain the measured oxygen saturation as measured by pulse oximetry > 92% or to prevent or reverse signs of inadequate ventilation during positive-pressure mask ventilation under GA		
Flow and timing	No information given		
Comparative			
Notes			
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?			
Was a case-control design avoided?			
Did the study avoid inappropriate exclusions?			
DOMAIN 2: Index Test All Tests			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			

Langeron 2000 (*Continued*)

Did the assessors of the index test have appropriate training?			
Was interobserver variability reported for some or all patients?			
Was interobserver agreement acceptable?			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?			
Were the reference standard results interpreted without knowledge of the results of the index tests?			
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?			
Did all patients receive the same reference standard?			
Were all patients included in the analysis?			

Lee 2015

Study characteristics	
Patient sampling	The 2011 year operating schedule was reviewed to identify study patients
Patient characteristics and setting	Exclusion criteria were: children, nasotracheal intubation, emergency intubation, fiberoptic-assisted intubation, existing tracheostomies or laryngectomies, laryngeal mask airway cases, regional anaesthesia without intubation, and incomplete charts. Inclusion criteria were adult (18 years) male and female patients undergoing direct laryngoscopy for the purpose of general

Lee 2015 (*Continued*)

	endotracheal anaesthesia Sample size: 344		
Index tests	MMT		
Target condition and reference standard(s)	Difficult laryngoscopy: Cormack and Lehane		
Flow and timing	Not stated in study		
Comparative			
Notes			
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?			
Was a case-control design avoided?			
Did the study avoid inappropriate exclusions?			
DOMAIN 2: Index Test All Tests			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Did the assessors of the index test have appropriate training?			
Was interobserver variability reported for some or all patients?			
Was interobserver agreement acceptable?			

Lee 2015 (*Continued*)

DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?			
Were the reference standard results interpreted without knowledge of the results of the index tests?			
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?			
Did all patients receive the same reference standard?			
Were all patients included in the analysis?			

Liaskou 2014

Study characteristics	
Patient sampling	387 consecutive adult patients (age > 18 years) ASA I-II, without known airway pathology, scheduled for surgical procedures under GA with tracheal intubation were assessed for eligibility
Patient characteristics and setting	Teaching hospital, patients scheduled for surgery Sample size: 341 178 females Mean age: 50 years
Index tests	SMD (< 15 cm)
Target condition and reference standard(s)	Difficult laryngoscopy: Cormack and Lehane
Flow and timing	Not described

Comparative			
Notes			
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?			
Was a case-control design avoided?			
Did the study avoid inappropriate exclusions?			
DOMAIN 2: Index Test All Tests			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Did the assessors of the index test have appropriate training?			
Was interobserver variability reported for some or all patients?			
Was interobserver agreement acceptable?			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?			

Were the reference standard results interpreted without knowledge of the results of the index tests?			
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?			
Did all patients receive the same reference standard?			
Were all patients included in the analysis?			

Lundstrom 2009

Study characteristics			
Patient sampling	Nationwide prospective registry		
Patient characteristics and setting	For this retrospective analysis patients with regional anaesthesia, sedation alone, no planned endotracheal intubation, intubation previous to OP, fibre optic intubation were excluded Sample size: 103,728 59,287 females		
Index tests	MMT		
Target condition and reference standard(s)	Difficult tracheal intubation: more than two attempts or more than one anaesthesist		
Flow and timing	Unclear		
Comparative			
Notes			
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			

Was a consecutive or random sample of patients enrolled?			
Was a case-control design avoided?			
Did the study avoid inappropriate exclusions?			
DOMAIN 2: Index Test All Tests			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Did the assessors of the index test have appropriate training?			
Was interobserver variability reported for some or all patients?			
Was interobserver agreement acceptable?			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?			
Were the reference standard results interpreted without knowledge of the results of the index tests?			
DOMAIN 4: Flow and Timing			

Lundstrom 2009 (*Continued*)

Was there an appropriate interval between index test and reference standard?			
Did all patients receive the same reference standard?			
Were all patients included in the analysis?			

Mallat 2010

Study characteristics			
Patient sampling	Patients were selected when at least one of the following criteria was found at preoperative evaluation: inability to palpate the cricoid cartilage, endothoracic goitre (every goitre that extends below the manubrium on the chest x-ray), tracheal deviation of more than 1 cm or tracheal stenosis on the chest x-ray		
Patient characteristics and setting	Patients with goitre only (see above) Sample size: 80 Mean age: 56 years		
Index tests	MMT		
Target condition and reference standard(s)	Difficult tracheal intubation: IDS		
Flow and timing	No details given		
Comparative			
Notes			
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?			
Was a case-control design avoided?			

Did the study avoid inappropriate exclusions?			
DOMAIN 2: Index Test All Tests			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Did the assessors of the index test have appropriate training?			
Was interobserver variability reported for some or all patients?			
Was interobserver agreement acceptable?			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?			
Were the reference standard results interpreted without knowledge of the results of the index tests?			
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?			
Did all patients receive the same reference standard?			

Were all patients included in the analysis?			

Mashour 2008

Study characteristics			
Patient sampling	All patients receiving anaesthesia with BMI >= 40		
Patient characteristics and setting	Only patients with BMI >= 40 Sample size: 346 231 females Mean age: 50 years		
Index tests	MMT, combination of tests		
Target condition and reference standard(s)	Difficult laryngoscopy: Cormack and Lehane		
Flow and timing	No information given		
Comparative			
Notes			
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?			
Was a case-control design avoided?			
Did the study avoid inappropriate exclusions?			
DOMAIN 2: Index Test All Tests			

Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Did the assessors of the index test have appropriate training?			
Was interobserver variability reported for some or all patients?			
Was interobserver agreement acceptable?			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?			
Were the reference standard results interpreted without knowledge of the results of the index tests?			
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?			
Did all patients receive the same reference standard?			
Were all patients included in the analysis?			

Mehta 2014

Study characteristics			
Patient sampling	Patients posted for elective surgery under GA Patients with obvious head and neck pathology, edentulous patients, mass in the mouth, BMI > 40, protruding upper incisors (total of 34) were excluded from the study		
Patient characteristics and setting	Sample size: 484 130 females Mean age: 44 years		
Index tests	MMT, TMD (< 6 cm), SMD (< 1.5 cm), mouth opening, ULBT		
Target condition and reference standard(s)	Difficult laryngoscopy: Cormack and Lehane		
Flow and timing	No information		
Comparative			
Notes			
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?			
Was a case-control design avoided?			
Did the study avoid inappropriate exclusions?			
DOMAIN 2: Index Test All Tests			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			

Mehta 2014 (*Continued*)

Did the assessors of the index test have appropriate training?			
Was interobserver variability reported for some or all patients?			
Was interobserver agreement acceptable?			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?			
Were the reference standard results interpreted without knowledge of the results of the index tests?			
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?			
Did all patients receive the same reference standard?			
Were all patients included in the analysis?			

Merah 2004

Study characteristics	
Patient sampling	ASA I-III patients selected for GA for caesarean section
Patient characteristics and setting	Exclusion: inability to sit, gross anatomical abnormity of head and neck, recent surgery of this areas, severe cardiorespiratory disorders Sample size: 80 80 females

Merah 2004 (*Continued*)

	Mean age: 30.9 years		
Index tests	MMT, TMD, SMD (< 13.5 cm), mouth opening (< 2.5 cm), combination of tests		
Target condition and reference standard(s)	Difficult laryngoscopy: Cormack and Lehane		
Flow and timing	Unclear		
Comparative			
Notes			
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?			
Was a case-control design avoided?			
Did the study avoid inappropriate exclusions?			
DOMAIN 2: Index Test All Tests			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Did the assessors of the index test have appropriate training?			
Was interobserver variability reported for some or all patients?			
Was interobserver agreement acceptable?			

DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?			
Were the reference standard results interpreted without knowledge of the results of the index tests?			
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?			
Did all patients receive the same reference standard?			
Were all patients included in the analysis?			

Mishra 2009

Study characteristics	
Patient sampling	No details given
Patient characteristics and setting	100 pregnant patients posted for caesarean section under GA (both emergency and elective) Patients with a history of burns, trauma, tumours or a mass and previous surgery involving the craniofaciocervical region or the airway, patients with restricted mobility of the neck andmandible (e.g. rheumatoid arthritis or cervical disk disorders), and severe pregnancy induced hypertension were excluded from the study Sample size: 100 100 females Mean age: 25 years
Index tests	MMT, ULBT
Target condition and reference standard(s)	Difficult laryngoscopy: Cormack and Lehane

Mishra 2009 (*Continued*)

Flow and timing	Unclear		
Comparative			
Notes			
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?			
Was a case-control design avoided?			
Did the study avoid inappropriate exclusions?			
DOMAIN 2: Index Test All Tests			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Did the assessors of the index test have appropriate training?			
Was interobserver variability reported for some or all patients?			
Was interobserver agreement acceptable?			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?			

Mishra 2009 (*Continued*)

Were the reference standard results interpreted without knowledge of the results of the index tests?			
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?			
Did all patients receive the same reference standard?			
Were all patients included in the analysis?			

Montemayor-Cruz 2015

Study characteristics	
Patient sampling	A non-probabilistic sample was performed by selecting consecutive cases over the month of January 2014
Patient characteristics and setting	Inclusion criteria: male and female patients of 15 to 75 years of age;elective surgical procedure; GA requiring orotracheal intubation Exclusion criteria: patients who refused to participate in the study (in the case of minors, patients whose legal guardians refused their participation in the study); patients who, due to their clinical status, were unable to co-operate with airway assessment (low Glasgow Coma Scale score, mental retardation, dementia, etc.); anatomical abnormalities altering the airway (deformity, tumours, etc.) and that precluded airway exploration regardless of the diagnosis the surgical procedure was to be performed for; patients already intubated Sample size: 70 35 females Mean age: 48 years
Index tests	MMT, mouth opening, combination of tests
Target condition and reference standard(s)	Difficult laryngoscopy: Cormack and Lehane
Flow and timing	Not stated
Comparative	

Notes			
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?			
Was a case-control design avoided?			
Did the study avoid inappropriate exclusions?			
DOMAIN 2: Index Test All Tests			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Did the assessors of the index test have appropriate training?			
Was interobserver variability reported for some or all patients?			
Was interobserver agreement acceptable?			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?			
Were the reference standard results interpreted without knowledge			

of the results of the index tests?			
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?			
Did all patients receive the same reference standard?			
Were all patients included in the analysis?			

Myneni 2010

Study characteristics			
Patient sampling	"All adult patients 18 years of age and older, presenting in all surgical specialties, were included in the study except for obstetric anesthesia or plastic surgery for burns"		
Patient characteristics and setting	Sample size: 6882		
Index tests	ULBT		
Target condition and reference standard(s)	Difficult laryngoscopy: Cormack and Lehane		
Flow and timing	Unclear		
Comparative			
Notes			
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?			

Was a case-control design avoided?			
Did the study avoid inappropriate exclusions?			
DOMAIN 2: Index Test All Tests			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Did the assessors of the index test have appropriate training?			
Was interobserver variability reported for some or all patients?			
Was interobserver agreement acceptable?			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?			
Were the reference standard results interpreted without knowledge of the results of the index tests?			
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?			

Did all patients receive the same reference standard?			
Were all patients included in the analysis?			

Nadal 1998

Study characteristics			
Patient sampling	All diabetic patients for elective surgery under GA included from May 1994 to May 1995		
Patient characteristics and setting	Excluded: obvious anatomical variations of face, neck, palate or hands, or had a history of difficult tracheal intubation Sample size: 83 53 females Mean age: 53 years		
Index tests	MT, TMD (< 6 cm)		
Target condition and reference standard(s)	Difficult laryngoscopy: Cormack and Lehane		
Flow and timing	Index test done one day before surgery		
Comparative			
Notes			
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?			
Was a case-control design avoided?			
Did the study avoid inappropriate exclusions?			

Nadal 1998 (*Continued*)

DOMAIN 2: Index Test All Tests			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Did the assessors of the index test have appropriate training?			
Was interobserver variability reported for some or all patients?			
Was interobserver agreement acceptable?			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?			
Were the reference standard results interpreted without knowledge of the results of the index tests?			
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?			
Did all patients receive the same reference standard?			
Were all patients included in the analysis?			

Naguib 1999

Study characteristics			
Patient sampling	Case-control		
Patient characteristics and setting	Patients in whom an unanticipated difficult intubation was identified and were scheduled to undergo endotracheal anaesthesia for any type of non-emergency surgical procedures except traumatic facial abnormalities or obstetric and cardiac surgery. Also random control group whom laryngoscopy and intubation was found to be easy and anaesthetized by the same anaesthesiologists Sample size: 57 15 females Mean age: 36.9 years		
Index tests	MMT		
Target condition and reference standard(s)	Difficult laryngoscopy: Cormack and Lehane		
Flow and timing	No information given		
Comparative			
Notes			
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?			
Was a case-control design avoided?			
Did the study avoid inappropriate exclusions?			
DOMAIN 2: Index Test All Tests			
Were the index test results interpreted without knowledge of the results of the reference standard?			

Naguib 1999 (*Continued*)

If a threshold was used, was it pre-specified?			
Did the assessors of the index test have appropriate training?			
Was interobserver variability reported for some or all patients?			
Was interobserver agreement acceptable?			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?			
Were the reference standard results interpreted without knowledge of the results of the index tests?			
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?			
Did all patients receive the same reference standard?			
Were all patients included in the analysis?			

Naguib 2006

Study characteristics	
Patient sampling	Case-control

Naguib 2006 (*Continued*)

Patient characteristics and setting	Adult patients presenting for GA for any type of non-emergency surgical procedures except traumatic facial abnormalities, obstetric surgery, or cardiac surgery with unanticipated difficult intubation. Also second patient from same day as control Sample size: 194 84 females Mean age: 53.7 years
Index tests	MMT
Target condition and reference standard(s)	Difficult tracheal intubation: two or more attempts at placing the endotracheal tube or the use of an alternative device
Flow and timing	No information given
Comparative	
Notes	

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?			
Was a case-control design avoided?			
Did the study avoid inappropriate exclusions?			
DOMAIN 2: Index Test All Tests			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Did the assessors of the index test have appropriate training?			

Naguib 2006 (*Continued*)

Was interobserver variability reported for some or all patients?			
Was interobserver agreement acceptable?			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?			
Were the reference standard results interpreted without knowledge of the results of the index tests?			
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?			
Did all patients receive the same reference standard?			
Were all patients included in the analysis?			

Nasir 2011

Study characteristics	
Patient sampling	122 patients were selected from the operative schedule by convenient non-probability sampling
Patient characteristics and setting	Adult patients belonging to ASA-I , II and III ranging from 18-65 years of either gender undergoing elective procedures from all surgical specialties requiring endotracheal intubation were included in the study. Emergency surgical procedures, patients with age < 18 years, pregnant patients, patients with unstable cervical spine and patients with tumour of the larynx were excluded Sample size: 122 79 females Mean age: 32.8 years

Nasir 2011 (*Continued*)

Index tests	MMT		
Target condition and reference standard(s)	Difficult laryngoscopy: Cormack and Lehane		
Flow and timing	Unknown		
Comparative			
Notes			
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?			
Was a case-control design avoided?			
Did the study avoid inappropriate exclusions?			
DOMAIN 2: Index Test All Tests			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Did the assessors of the index test have appropriate training?			
Was interobserver variability reported for some or all patients?			
Was interobserver agreement acceptable?			

Nasir 2011 (*Continued*)

DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?			
Were the reference standard results interpreted without knowledge of the results of the index tests?			
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?			
Did all patients receive the same reference standard?			
Were all patients included in the analysis?			

Nasiri 2013

Study characteristics	
Patient sampling	"Our study population included all patients who were referred for elective surgery, adult patients aged 18 to 75 years. Patients with burns, neck, tumors, head and neck injury were excluded"
Patient characteristics and setting	Sample size: 410 204 females
Index tests	Mouth opening, ULBT
Target condition and reference standard(s)	Difficult laryngoscopy: Cormack and Lehane
Flow and timing	Not reported
Comparative	
Notes	

Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?			
Was a case-control design avoided?			
Did the study avoid inappropriate exclusions?			
DOMAIN 2: Index Test All Tests			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Did the assessors of the index test have appropriate training?			
Was interobserver variability reported for some or all patients?			
Was interobserver agreement acceptable?			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?			
Were the reference standard results interpreted without knowledge of the results of the index tests?			

Nasiri 2013 (*Continued*)

DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?			
Did all patients receive the same reference standard?			
Were all patients included in the analysis?			

Nath 1997

Study characteristics			
Patient sampling	Case-control		
Patient characteristics and setting	Adults requiring GA and intubation (including easy and difficult intubations). Also 16 patients reported to be difficult to intubate Sample size: 300 127 females Mean age: 39.7 years		
Index tests	MMT		
Target condition and reference standard(s)	Difficult laryngoscopy: Cormack and Lehane		
Flow and timing	Index test postoperative. Reference standard was re-checked according to documentation for those who were included retrospectively		
Comparative			
Notes			
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?			

Was a case-control design avoided?			
Did the study avoid inappropriate exclusions?			
DOMAIN 2: Index Test All Tests			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Did the assessors of the index test have appropriate training?			
Was interobserver variability reported for some or all patients?			
Was interobserver agreement acceptable?			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?			
Were the reference standard results interpreted without knowledge of the results of the index tests?			
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?			

Did all patients receive the same reference standard?			
Were all patients included in the analysis?			

Noorizad 2006

Study characteristics			
Patient sampling	No details given		
Patient characteristics and setting	Sample size: 379 200 females		
Index tests	MMT, TMD (< 6 cm)		
Target condition and reference standard(s)	Difficult laryngoscopy: Cormack and Lehane		
Flow and timing	Index test at preoperative visit. Reference standard at OP		
Comparative			
Notes			
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?			
Was a case-control design avoided?			
Did the study avoid inappropriate exclusions?			
DOMAIN 2: Index Test All Tests			

Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Did the assessors of the index test have appropriate training?			
Was interobserver variability reported for some or all patients?			
Was interobserver agreement acceptable?			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?			
Were the reference standard results interpreted without knowledge of the results of the index tests?			
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?			
Did all patients receive the same reference standard?			
Were all patients included in the analysis?			

Oates 1990

Study characteristics			
Patient sampling	Subgroup of patients scheduled for elective surgery from an unpublished prospective study		
Patient characteristics and setting	Sample size: 334		
Index tests	MT		
Target condition and reference standard(s)	Difficult laryngoscopy: Cormack and Lehane		
Flow and timing	No information given		
Comparative			
Notes			
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?			
Was a case-control design avoided?			
Did the study avoid inappropriate exclusions?			
DOMAIN 2: Index Test All Tests			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Did the assessors of the index test have appropriate training?			

Oates 1990 (*Continued*)

Was interobserver variability reported for some or all patients?			
Was interobserver agreement acceptable?			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?			
Were the reference standard results interpreted without knowledge of the results of the index tests?			
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?			
Did all patients receive the same reference standard?			
Were all patients included in the analysis?			

Oates 1991

Study characteristics	
Patient sampling	Patients requiring tracheal intubation for operation. No further details
Patient characteristics and setting	Sample size: 751 448 females
Index tests	MT, Wilson risk score
Target condition and reference standard(s)	Difficult laryngoscopy: Cormack and Lehane

Oates 1991 (*Continued*)

Flow and timing	No information given		
Comparative			
Notes			
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?			
Was a case-control design avoided?			
Did the study avoid inappropriate exclusions?			
DOMAIN 2: Index Test All Tests			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Did the assessors of the index test have appropriate training?			
Was interobserver variability reported for some or all patients?			
Was interobserver agreement acceptable?			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?			

Oates 1991 (*Continued*)

Were the reference standard results interpreted without knowledge of the results of the index tests?			
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?			
Did all patients receive the same reference standard?			
Were all patients included in the analysis?			

Pottecher 1991

Study characteristics			
Patient sampling	Gynaecologic-obstetric patients requiring intubation for OP		
Patient characteristics and setting	Sample size: 663 663 females Mean age: 37.9 years		
Index tests	MMT (I versus II-IV), TMD (< 8 cm), SMD, mouth opening (< 4.1 cm)		
Target condition and reference standard(s)	Difficult laryngoscopy: Cormack and Lehane; difficult tracheal intubation		
Flow and timing	Unclear		
Comparative			
Notes			
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			

Pottecher 1991 (*Continued*)

Was a consecutive or random sample of patients enrolled?			
Was a case-control design avoided?			
Did the study avoid inappropriate exclusions?			
DOMAIN 2: Index Test All Tests			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Did the assessors of the index test have appropriate training?			
Was interobserver variability reported for some or all patients?			
Was interobserver agreement acceptable?			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?			
Were the reference standard results interpreted without knowledge of the results of the index tests?			
DOMAIN 4: Flow and Timing			

Pottecher 1991 (*Continued*)

Was there an appropriate interval between index test and reference standard?			
Did all patients receive the same reference standard?			
Were all patients included in the analysis?			

Prakash 2013

Study characteristics			
Patient sampling	"Patients under general anaesthesia requiring tracheal intubation were included in this prospective study"		
Patient characteristics and setting	Adult ASA I and II adult patients scheduled for elective surgery. Patients with obvious abnormality of the airway where intubation under GA would be contraindicated, those at increased risk of aspiration, inter-incisor distance < 2.5 cm and unstable cervical spine were excluded from the study Sample size: 330 222 females Mean age: 37.8 years		
Index tests	MMT, mouth opening, combination of tests		
Target condition and reference standard(s)	Difficult laryngoscopy: Cormack and Lehane		
Flow and timing	Not provided		
Comparative			
Notes			
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?			
Was a case-control design avoided?			

Did the study avoid inappropriate exclusions?			
DOMAIN 2: Index Test All Tests			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Did the assessors of the index test have appropriate training?			
Was interobserver variability reported for some or all patients?			
Was interobserver agreement acceptable?			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?			
Were the reference standard results interpreted without knowledge of the results of the index tests?			
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?			
Did all patients receive the same reference standard?			

Were all patients included in the analysis?			

Qudaisat 2011

Study characteristics			
Patient sampling	Unclear. "variety of elective procedures under general anaesthesia" Exclusion: facial asymmetry, upper incisor protrusion, edentulousness, limited mouth opening		
Patient characteristics and setting	Sample size: 235 98 females		
Index tests	TMD		
Target condition and reference standard(s)	Difficult laryngoscopy: Cormack and Lehane		
Flow and timing	Unclear		
Comparative			
Notes			
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?			
Was a case-control design avoided?			
Did the study avoid inappropriate exclusions?			
DOMAIN 2: Index Test All Tests			
Were the index test results interpreted without knowledge of the results of the reference stan-			

Qudaisat 2011 (*Continued*)

dard?			
If a threshold was used, was it pre-specified?			
Did the assessors of the index test have appropriate training?			
Was interobserver variability reported for some or all patients?			
Was interobserver agreement acceptable?			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?			
Were the reference standard results interpreted without knowledge of the results of the index tests?			
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?			
Did all patients receive the same reference standard?			
Were all patients included in the analysis?			

Reghunathan 2016

Study characteristics			
Patient sampling	Patients of both sexes, between 15 and 55 years, and belonging to ASA grade I or II were selected. Patients with obesity, malposition of teeth, microstomia, macroglossia, edentulous or with artificial dentures, cervical spondylosis, short neck, contractures of neck, neck swellings, postradiation fibrosis, developmental anomalies which may affect airway assessment, and in whom difficult intubation was expected were excluded from the study		
Patient characteristics and setting	Sample size: 200 Mean age: 35 years		
Index tests	Combination of tests		
Target condition and reference standard(s)	Difficult laryngoscopy: Cormack and Lehane		
Flow and timing	Unclear		
Comparative			
Notes			
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?			
Was a case-control design avoided?			
Did the study avoid inappropriate exclusions?			
DOMAIN 2: Index Test All Tests			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			

Reghunathan 2016 (*Continued*)

Did the assessors of the index test have appropriate training?			
Was interobserver variability reported for some or all patients?			
Was interobserver agreement acceptable?			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?			
Were the reference standard results interpreted without knowledge of the results of the index tests?			
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?			
Did all patients receive the same reference standard?			
Were all patients included in the analysis?			

Rocke 1992

Study characteristics	
Patient sampling	All patients undergoing elective or emergency caesarean section under GA; no further information
Patient characteristics and setting	Sample size: 1500 1500 females Mean age: 26.4 years

Rocke 1992 (*Continued*)

Index tests	MMT		
Target condition and reference standard(s)	Difficult laryngoscopy: Cormack and Lehane; difficult tracheal intubation		
Flow and timing	No information		
Comparative			
Notes			
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?			
Was a case-control design avoided?			
Did the study avoid inappropriate exclusions?			
DOMAIN 2: Index Test All Tests			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Did the assessors of the index test have appropriate training?			
Was interobserver variability reported for some or all patients?			
Was interobserver agreement acceptable?			

Rocke 1992 (*Continued*)

DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?			
Were the reference standard results interpreted without knowledge of the results of the index tests?			
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?			
Did all patients receive the same reference standard?			
Were all patients included in the analysis?			

Safavi 2014

Study characteristics	
Patient sampling	Consecutive ASA I-III adult patients
Patient characteristics and setting	"These patients were scheduled for elective surgery under general anesthesia requiring endotracheal intubation" Sample size: 476 150 females Mean age: 36.6 years
Index tests	MMT, ULBT, combination of tests
Target condition and reference standard(s)	Difficult laryngoscopy: Cormack and Lehane
Flow and timing	Not described
Comparative	

Notes			
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?			
Was a case-control design avoided?			
Did the study avoid inappropriate exclusions?			
DOMAIN 2: Index Test All Tests			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Did the assessors of the index test have appropriate training?			
Was interobserver variability reported for some or all patients?			
Was interobserver agreement acceptable?			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?			
Were the reference standard results interpreted without knowledge			

Safavi 2014 (*Continued*)

of the results of the index tests?			
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?			
Did all patients receive the same reference standard?			
Were all patients included in the analysis?			

Sahin 2011

Study characteristics			
Patient sampling	ASA I-III patients scheduled for elective surgery and requiring endotracheal intubation were enrolled in the study prospectively over a 1-year period		
Patient characteristics and setting	The exclusion criteria included patients aged less than 18 years, obstetric patients, patients with anatomic abnormality or recent surgery of the head/neck, burns or trauma to the airways or to the cranial, cervical, and facial regions, patients with tumours or a mass in the aforementioned regions, patients with restricted mobility of the neck and mandible, and patients who do not have incisor teeth Sample size: 762 367 females		
Index tests	Combination of tests		
Target condition and reference standard(s)	Difficult tracheal intubation		
Flow and timing	Index tests: during the preoperative visit Reference standard: after induction of GA		
Comparative			
Notes			
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**

Sahin 2011 (*Continued*)

DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?			
Was a case-control design avoided?			
Did the study avoid inappropriate exclusions?			
DOMAIN 2: Index Test All Tests			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Did the assessors of the index test have appropriate training?			
Was interobserver variability reported for some or all patients?			
Was interobserver agreement acceptable?			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?			
Were the reference standard results interpreted without knowledge of the results of the index tests?			
DOMAIN 4: Flow and Timing			

Sahin 2011 (*Continued*)

Was there an appropriate interval between index test and reference standard?			
Did all patients receive the same reference standard?			
Were all patients included in the analysis?			

Salimi 2008

Study characteristics			
Patient sampling	All patients aged above 16 who required GA with endotracheal intubation for elective surgery within 1 year were included		
Patient characteristics and setting	Patients with a history of previous surgery, burns or trauma to the airways or to the cranial, cervical and facial regions, patients with tumours or a mass in the above-mentioned regions, patients with restricted motility of the neck and mandible (e.g. rheumatoid arthritis or cervical disk disorders) , patients without teeth, and patients with incomplete information forms were excluded from the study Sample size: 350 114 females Mean age: 32 years		
Index tests	TMD (< 4 cm), ULBT		
Target condition and reference standard(s)	Difficult laryngoscopy: Cormack and Lehane		
Flow and timing	Unclear		
Comparative			
Notes			
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?			

Was a case-control design avoided?			
Did the study avoid inappropriate exclusions?			
DOMAIN 2: Index Test All Tests			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Did the assessors of the index test have appropriate training?			
Was interobserver variability reported for some or all patients?			
Was interobserver agreement acceptable?			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?			
Were the reference standard results interpreted without knowledge of the results of the index tests?			
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?			

Did all patients receive the same reference standard?			
Were all patients included in the analysis?			

Samra 1995

Study characteristics			
Patient sampling	Random selection, no further details given		
Patient characteristics and setting	Patients with clinically obvious mandibular abnormalities (i.e. receding mandible, poor mobility of temporomandibular joint either due to arthritis, pain, trauma, or trismus) and those patients with history of arthritis and/or limitation of movement of cervical spine were excluded Sample size: 564		
Index tests	MMT		
Target condition and reference standard(s)	Difficult laryngoscopy: Cormack and Lehane		
Flow and timing	Unclear		
Comparative			
Notes			
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?			
Was a case-control design avoided?			
Did the study avoid inappropriate exclusions?			
DOMAIN 2: Index Test All Tests			

Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Did the assessors of the index test have appropriate training?			
Was interobserver variability reported for some or all patients?			
Was interobserver agreement acceptable?			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?			
Were the reference standard results interpreted without knowledge of the results of the index tests?			
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?			
Did all patients receive the same reference standard?			
Were all patients included in the analysis?			

Savva 1994

Study characteristics			
Patient sampling	Consecutive patients (322 of them obstetric) requiring tracheal intubation as part of their anaesthesia		
Patient characteristics and setting	Sample size: 350 185 females Mean age: 39 years		
Index tests	MMT, TMD, SMD, combination of tests		
Target condition and reference standard(s)	Difficult laryngoscopy: Cormack and Lehane		
Flow and timing	No information given		
Comparative			
Notes			
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?			
Was a case-control design avoided?			
Did the study avoid inappropriate exclusions?			
DOMAIN 2: Index Test All Tests			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Did the assessors of the index test have appropriate training?			

Savva 1994 (*Continued*)

Was interobserver variability reported for some or all patients?			
Was interobserver agreement acceptable?			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?			
Were the reference standard results interpreted without knowledge of the results of the index tests?			
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?			
Did all patients receive the same reference standard?			
Were all patients included in the analysis?			

Schmitt 2000

Study characteristics	
Patient sampling	"Between March 1994 and December 1998, all acromegalic patients (American Society of Anesthesiologists class I-III,68 women, 60 men) scheduled for elective transsphenoidal resection of a growth hormone secreting pituitary adenoma were investigated. The diagnosis of acromegaly was contirmed by clinical and endocrine reassessment (failure to suppress growth hormone to < 2 pg/l after an oral glucose load) as well as by magnetic resonance imaging showing the size and the extent of a pituitary adenoma just before surgery"
Patient characteristics and setting	"All patients showed typical acromegalic features such as macro-glossia, prognathism, or soft tissue swelling in various degrees. Preoperatively, Mallampati classification, thyromental distance, and

Schmitt 2000 (*Continued*)

	head and neck movement were determined in each patient. After induction of anesthesia and muscle paralysis, laryngoscopic grade was assessed during direct laryngoscopy" Sample size: 128 68 females Mean age: 46 years		
Index tests	MMT, combination of tests		
Target condition and reference standard(s)	Difficult laryngoscopy: Cormack and Lehane		
Flow and timing	Preoperatively		
Comparative			
Notes			
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?			
Was a case-control design avoided?			
Did the study avoid inappropriate exclusions?			
DOMAIN 2: Index Test All Tests			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Did the assessors of the index test have appropriate training?			
Was interobserver variability reported for some or all patients?			

Schmitt 2000 (*Continued*)

Was interobserver agreement acceptable?			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?			
Were the reference standard results interpreted without knowledge of the results of the index tests?			
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?			
Did all patients receive the same reference standard?			
Were all patients included in the analysis?			

Seo 2012

Study characteristics	
Patient sampling	The study was performed at the hospital on 305 ASA I and II patients between ages 19 and 70 years, who were scheduled for surgery under GA
Patient characteristics and setting	Patients were excluded from the study if their teeth were incomplete, if the patient had limited head and neck movement, had impairment of the temporomandibular joint, or had oral or laryngeal tumour Sample size: 305 157 females
Index tests	MMT, TMD (< 6 cm), mouth opening, ULBT, combination of tests

Seo 2012 (*Continued*)

Target condition and reference standard(s)	Difficult laryngoscopy: Cormack and Lehane; difficult tracheal intubation: IDS		
Flow and timing	Unclear		
Comparative			
Notes			
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?			
Was a case-control design avoided?			
Did the study avoid inappropriate exclusions?			
DOMAIN 2: Index Test All Tests			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Did the assessors of the index test have appropriate training?			
Was interobserver variability reported for some or all patients?			
Was interobserver agreement acceptable?			
DOMAIN 3: Reference Standard			

Seo 2012 (*Continued*)

Is the reference standards likely to correctly classify the target condition?			
Were the reference standard results interpreted without knowledge of the results of the index tests?			
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?			
Did all patients receive the same reference standard?			
Were all patients included in the analysis?			

Shah 2013

Study characteristics	
Patient sampling	Adult patients of more than 18 years age, of either sex, of ASA grade I and II, undergoing elective surgeries under GA
Patient characteristics and setting	Patients unable to sit or stand erect, pregnant females, those having obvious malformation of the airway or those requiring awake intubation were excluded from the study Sample size: 480 241 females
Index tests	MMT, TMD, mouth opening (< 4 cm), ULBT, combination of tests
Target condition and reference standard(s)	Difficult laryngoscopy: Cormack and Lehane
Flow and timing	Not provided
Comparative	
Notes	

Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?			
Was a case-control design avoided?			
Did the study avoid inappropriate exclusions?			
DOMAIN 2: Index Test All Tests			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Did the assessors of the index test have appropriate training?			
Was interobserver variability reported for some or all patients?			
Was interobserver agreement acceptable?			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?			
Were the reference standard results interpreted without knowledge of the results of the index tests?			

DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?			
Did all patients receive the same reference standard?			
Were all patients included in the analysis?			

Shah 2014

Study characteristics			
Patient sampling	Patients aged ≥16 years, scheduled to undergo surgery under GA were included in the study using nonprobability consecutive sampling		
Patient characteristics and setting	Edentulous patients, those unable to open the mouth, those with laryngeal masses, those having large goiters or with limitation of cervical movements were excluded from the study Sample size: 450 254 females Mean ag: 38.8 years		
Index tests	ULBT		
Target condition and reference standard(s)	Difficult tracheal intubation		
Flow and timing	Preoperatively		
Comparative			
Notes			
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?			

Was a case-control design avoided?			
Did the study avoid inappropriate exclusions?			
DOMAIN 2: Index Test All Tests			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Did the assessors of the index test have appropriate training?			
Was interobserver variability reported for some or all patients?			
Was interobserver agreement acceptable?			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?			
Were the reference standard results interpreted without knowledge of the results of the index tests?			
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?			

Did all patients receive the same reference standard?			
Were all patients included in the analysis?			

Sharma 2010

Study characteristics			
Patient sampling	Case-control		
Patient characteristics and setting	Over a period of 5 years, 64 consecutive acromegalic patients presenting for surgery for excision of pituitary tumor were enrolled. For each acromegaly patient enrolled, the subsequent nonacromegalic patient scheduled for excision of pituitary tumour during the same 5-year period was also enrolled to serve as a control Sample size: 125		
Index tests	MMT, ULBT		
Target condition and reference standard(s)	Difficult laryngoscopy: Cormack and Lehane		
Flow and timing	Unclear		
Comparative			
Notes			
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?			
Was a case-control design avoided?			
Did the study avoid inappropriate exclusions?			

DOMAIN 2: Index Test All Tests			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Did the assessors of the index test have appropriate training?			
Was interobserver variability reported for some or all patients?			
Was interobserver agreement acceptable?			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?			
Were the reference standard results interpreted without knowledge of the results of the index tests?			
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?			
Did all patients receive the same reference standard?			
Were all patients included in the analysis?			

Singh 2009

Study characteristics			
Patient sampling	No information of selection process		
Patient characteristics and setting	ASA I and II patients undergoing elective lower segment caesarean section under GA. Women with full stomach and apparent abnormalities of the neck and face were excluded Sample size: 300		
Index tests	MMT (I versus II-IV), Wilson risk score, TMD (< 5 cm), combination of tests		
Target condition and reference standard(s)	Difficult laryngoscopy: Cormack and Lehane		
Flow and timing	Unclear		
Comparative			
Notes			
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?			
Was a case-control design avoided?			
Did the study avoid inappropriate exclusions?			
DOMAIN 2: Index Test All Tests			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Did the assessors of the index test have appropriate training?			

Singh 2009 (*Continued*)

Was interobserver variability reported for some or all patients?			
Was interobserver agreement acceptable?			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?			
Were the reference standard results interpreted without knowledge of the results of the index tests?			
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?			
Did all patients receive the same reference standard?			
Were all patients included in the analysis?			

Soyuncu 2009

Study characteristics	
Patient sampling	All patients who required intubation in the ED were included in the study
Patient characteristics and setting	ED patients Sample size: 366 115 females Mean age: 46.8 years
Index tests	Mouth opening (< 3 cm), combination of tests

Target condition and reference standard(s)	Difficult tracheal intubation		
Flow and timing	Unclear		
Comparative			
Notes			
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?			
Was a case-control design avoided?			
Did the study avoid inappropriate exclusions?			
DOMAIN 2: Index Test All Tests			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Did the assessors of the index test have appropriate training?			
Was interobserver variability reported for some or all patients?			
Was interobserver agreement acceptable?			
DOMAIN 3: Reference Standard			

Soyuncu 2009 (*Continued*)

Is the reference standards likely to correctly classify the target condition?			
Were the reference standard results interpreted without knowledge of the results of the index tests?			
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?			
Did all patients receive the same reference standard?			
Were all patients included in the analysis?			

Tantri 2016

Study characteristics	
Patient sampling	"Patients who underwent elective surgery with general anesthesia were included in this study"
Patient characteristics and setting	The inclusion criteria were patients aged 18 to 65 years old; an ASA score of 1 or 2; Indonesians of Malay race; and willingness to participate in this study, as indicated by signing the informed consent form. Patients with oral opening restricted to less than 3 cm, acute burns on the face and neck, tumours on the airway, limitations on neck movement, airway trauma, protruding upper teeth, a high risk of bleeding, acute respiratory infection (croup, epiglottitis, Ludwig's angina), or anatomical disturbances (macroglossia, short neck, micrognathia, prognathism) were excluded from this study Sample size: 277 160 females Mean age: 40.38 years
Index tests	MMT, TMD, combination of tests
Target condition and reference standard(s)	Difficult laryngoscopy: Cormack and Lehane
Flow and timing	Not described

Comparative			
Notes			
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?			
Was a case-control design avoided?			
Did the study avoid inappropriate exclusions?			
DOMAIN 2: Index Test All Tests			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Did the assessors of the index test have appropriate training?			
Was interobserver variability reported for some or all patients?			
Was interobserver agreement acceptable?			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?			

Tantri 2016 (*Continued*)

Were the reference standard results interpreted without knowledge of the results of the index tests?			
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?			
Did all patients receive the same reference standard?			
Were all patients included in the analysis?			

Thompson 2009

Study characteristics			
Patient sampling	Database of obstetric patients who underwent tracheal intubation and who had MMT and Cormack and Lehane grade recorded		
Patient characteristics and setting	Sample size: 1602		
Index tests	MMT		
Target condition and reference standard(s)	Difficult laryngoscopy: Cormack and Lehane		
Flow and timing	No information given		
Comparative			
Notes			
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			

Was a consecutive or random sample of patients enrolled?			
Was a case-control design avoided?			
Did the study avoid inappropriate exclusions?			
DOMAIN 2: Index Test All Tests			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Did the assessors of the index test have appropriate training?			
Was interobserver variability reported for some or all patients?			
Was interobserver agreement acceptable?			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?			
Were the reference standard results interpreted without knowledge of the results of the index tests?			
DOMAIN 4: Flow and Timing			

Thompson 2009 (*Continued*)

Was there an appropriate interval between index test and reference standard?			
Did all patients receive the same reference standard?			
Were all patients included in the analysis?			

Tse 1995

Study characteristics			
Patient sampling	Consecutive male and female patients aged 18 years and older undergoing elective surgery		
Patient characteristics and setting	Excluded patients with obvious malformations of airway, edentulous patiens, and patients who required cricoid pressure for RSI Sample size: 471 251 females		
Index tests	MT, TMD (< 7 cm), combination of tests		
Target condition and reference standard(s)	Difficult laryngoscopy: Cormack and Lehane		
Flow and timing	Preoperative measurements recorded on a form not seen by attending anaesthetist. Exact time interval not specified		
Comparative			
Notes			
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?			
Was a case-control design avoided?			

Tse 1995 (*Continued*)

Did the study avoid inappropriate exclusions?			
DOMAIN 2: Index Test All Tests			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Did the assessors of the index test have appropriate training?			
Was interobserver variability reported for some or all patients?			
Was interobserver agreement acceptable?			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?			
Were the reference standard results interpreted without knowledge of the results of the index tests?			
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?			
Did all patients receive the same reference standard?			

Tse 1995 (*Continued*)

Were all patients included in the analysis?			

Tuzuner-Oncul 2008

Study characteristics			
Patient sampling	No details on selection		
Patient characteristics and setting	Adult maxillofacial surgery patients Sample size: 208 108 females Mean age: 29 years		
Index tests	MMT, TMD (< 6 cm), SMD, mouth opening test (< 2.5 cm), combination of tests		
Target condition and reference standard(s)	Difficult tracheal intubation		
Flow and timing	Unclear		
Comparative			
Notes			
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?			
Was a case-control design avoided?			
Did the study avoid inappropriate exclusions?			
DOMAIN 2: Index Test All Tests			

Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Did the assessors of the index test have appropriate training?			
Was interobserver variability reported for some or all patients?			
Was interobserver agreement acceptable?			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?			
Were the reference standard results interpreted without knowledge of the results of the index tests?			
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?			
Did all patients receive the same reference standard?			
Were all patients included in the analysis?			

Ul Haq 2013

Study characteristics			
Patient sampling	Patients from the preoperative clinic, preoperative waiting area, and operating rooms were enrolled. ASA I-III patients aged above 18 years of either sex who were scheduled for elective surgeries under GA requiring tracheal intubation were included in the study		
Patient characteristics and setting	Sample size: 760 422 females Mean age: 43.44 years		
Index tests	MMT, combination of tests		
Target condition and reference standard(s)	Difficult laryngoscopy: Cormack and Lehane		
Flow and timing	Preoperative, operative, no time interval reported		
Comparative			
Notes			
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?			
Was a case-control design avoided?			
Did the study avoid inappropriate exclusions?			
DOMAIN 2: Index Test All Tests			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			

Ul Haq 2013 (*Continued*)

Did the assessors of the index test have appropriate training?			
Was interobserver variability reported for some or all patients?			
Was interobserver agreement acceptable?			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?			
Were the reference standard results interpreted without knowledge of the results of the index tests?			
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?			
Did all patients receive the same reference standard?			
Were all patients included in the analysis?			

Uribe 2015

Study characteristics	
Patient sampling	"A computerized search was initiated through the electronic medical records, which revealed 20,985 patients who underwent abdominal surgery requiring general anesthesia at The Ohio State University Wexner Medical Center during a period of 12 months, from January 1, 2007 to December 31, 2007. Using Microsoft Excel, every third patient from an alphabetized list was selected to generate a random sample of 6964 patients for this study. We performed a retrospective chart review of patients who underwent abdominal surgeries with ASA stratification I-V under general anesthesia

Uribe 2015 (*Continued*)

	requiring endotracheal intubation"		
Patient characteristics and setting	Sample size: 1970 2333 females		
Index tests	MMT		
Target condition and reference standard(s)	Difficult tracheal intubation		
Flow and timing	Not reported		
Comparative			
Notes			
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?			
Was a case-control design avoided?			
Did the study avoid inappropriate exclusions?			
DOMAIN 2: Index Test All Tests			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Did the assessors of the index test have appropriate training?			
Was interobserver variability reported for some or all patients?			

Uribe 2015 (*Continued*)

Was interobserver agreement acceptable?			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?			
Were the reference standard results interpreted without knowledge of the results of the index tests?			
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?			
Did all patients receive the same reference standard?			
Were all patients included in the analysis?			

Vallem 2015

Study characteristics	
Patient sampling	200 ASA grade I and II (18 to 60 years of age) adult patients scheduled to receive GA with endotracheal intubation
Patient characteristics and setting	Patients with airway malformations, edentulous patients, pregnancy and lactating mothers and patients with craniofacial anamolies were excluded from the study. Preoperative airway examination was performed using multiple screening tests to predict difficult airway Sample size: 200 35 females Mean age: 39.5 years
Index tests	MMT, TMD (< 6 cm), SMD (< 11 cm), mouth opening, ULBT

Target condition and reference standard(s)	Difficult laryngoscopy: Cormack and Lehane		
Flow and timing	Unclear		
Comparative			
Notes			
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?			
Was a case-control design avoided?			
Did the study avoid inappropriate exclusions?			
DOMAIN 2: Index Test All Tests			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Did the assessors of the index test have appropriate training?			
Was interobserver variability reported for some or all patients?			
Was interobserver agreement acceptable?			
DOMAIN 3: Reference Standard			

Vallem 2015 (*Continued*)

Is the reference standards likely to correctly classify the target condition?			
Were the reference standard results interpreted without knowledge of the results of the index tests?			
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?			
Did all patients receive the same reference standard?			
Were all patients included in the analysis?			

Vani 2000

Study characteristics	
Patient sampling	Patients with diabetes undergoing elective surgery
Patient characteristics and setting	Excluded: diabetes < 1 year, obvious anatomical malformation, history of difficult intubation Sample size: 50 28 females Mean age: 57.1 years
Index tests	MMT, TMD (< 6 cm)
Target condition and reference standard(s)	Difficult laryngoscopy: Cormack and Lehane
Flow and timing	Unclear
Comparative	
Notes	
Methodological quality	

Vani 2000 (*Continued*)

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?			
Was a case-control design avoided?			
Did the study avoid inappropriate exclusions?			
DOMAIN 2: Index Test All Tests			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Did the assessors of the index test have appropriate training?			
Was interobserver variability reported for some or all patients?			
Was interobserver agreement acceptable?			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?			
Were the reference standard results interpreted without knowledge of the results of the index tests?			

Vani 2000 (*Continued*)

DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?			
Did all patients receive the same reference standard?			
Were all patients included in the analysis?			

Wajekar 2015

Study characteristics			
Patient sampling	ASA I and II patients above 18 years undergoing elective surgical procedures requiring endotracheal intubation were included in the study		
Patient characteristics and setting	Patients with a history of previous surgery, burns or trauma, tumours/mass in the airways or the cranial, cervical and facial regions, patients with restricted mobility of the neck and mandible (rheumatoid arthritis, cervical disc disorders, or temporomandibular joint disorders), edentulous patients, pregnant patients, and BMI > 26 kg/m^2 were excluded from the study Sample size: 402 294 females Mean age: 41.9 years		
Index tests	MMT, TMD, ULBT		
Target condition and reference standard(s)	Difficult laryngoscopy: Cormack and Lehane		
Flow and timing	Unclear		
Comparative			
Notes			
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			

Was a consecutive or random sample of patients enrolled?			
Was a case-control design avoided?			
Did the study avoid inappropriate exclusions?			
DOMAIN 2: Index Test All Tests			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Did the assessors of the index test have appropriate training?			
Was interobserver variability reported for some or all patients?			
Was interobserver agreement acceptable?			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?			
Were the reference standard results interpreted without knowledge of the results of the index tests?			
DOMAIN 4: Flow and Timing			

Wajekar 2015 (*Continued*)

Was there an appropriate interval between index test and reference standard?			
Did all patients receive the same reference standard?			
Were all patients included in the analysis?			

Wilson 1988

Study characteristics			
Patient sampling	Patients > 16 years undergoing non-emergent surgery who underwent anaesthesia by four doctors		
Patient characteristics and setting	Sample size: 778		
Index tests	Wilson risk score		
Target condition and reference standard(s)	Difficult laryngoscopy: Cormack and Lehane		
Flow and timing	Index test was calculated retrospectively		
Comparative			
Notes			
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?			
Was a case-control design avoided?			
Did the study avoid inappropriate exclusions?			

DOMAIN 2: Index Test All Tests			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Did the assessors of the index test have appropriate training?			
Was interobserver variability reported for some or all patients?			
Was interobserver agreement acceptable?			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?			
Were the reference standard results interpreted without knowledge of the results of the index tests?			
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?			
Did all patients receive the same reference standard?			
Were all patients included in the analysis?			

Wong 1999

Study characteristics			
Patient sampling	All women scheduled for elective caesarean section under GA. Also all women scheduled for elective gynaecological OPs under GA		
Patient characteristics and setting	Sample size: 411 411 females Mean age: 27.9 years		
Index tests	MMT, TMD, combination of tests		
Target condition and reference standard(s)	Difficult laryngoscopy: Cormack and Lehane		
Flow and timing	Index test by primary invastigator. Reference standard by attending anaesthesiologist		
Comparative			
Notes			
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?			
Was a case-control design avoided?			
Did the study avoid inappropriate exclusions?			
DOMAIN 2: Index Test All Tests			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Did the assessors of the index test have appropriate training?			

Wong 1999 (*Continued*)

Was interobserver variability reported for some or all patients?			
Was interobserver agreement acceptable?			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?			
Were the reference standard results interpreted without knowledge of the results of the index tests?			
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?			
Did all patients receive the same reference standard?			
Were all patients included in the analysis?			

Wong 2009

Study characteristics	
Patient sampling	A prospective survey on consecutive adult patients scheduled for elective and emergency head and neck surgery requiring GA was performed. Data were collected over a 12-month period
Patient characteristics and setting	Sample size: 644 241 females Mean age: 52 years
Index tests	MMT, TMD, mouth opening (< 2.5 cm), combination of tests

Wong 2009 (*Continued*)

Target condition and reference standard(s)	Difficult laryngoscopy: Cormack and Lehane		
Flow and timing	Index tests during preoperative visit. Reference standard in the OR		
Comparative			
Notes			
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?			
Was a case-control design avoided?			
Did the study avoid inappropriate exclusions?			
DOMAIN 2: Index Test All Tests			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Did the assessors of the index test have appropriate training?			
Was interobserver variability reported for some or all patients?			
Was interobserver agreement acceptable?			
DOMAIN 3: Reference Standard			

Wong 2009 (*Continued*)

Is the reference standards likely to correctly classify the target condition?			
Were the reference standard results interpreted without knowledge of the results of the index tests?			
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?			
Did all patients receive the same reference standard?			
Were all patients included in the analysis?			

Yamamoto 1997

Study characteristics	
Patient sampling	Consecutive patients
Patient characteristics and setting	Routine patient care Sample size: 7270 3635 females Mean age: 52 years
Index tests	MMT, Wilson risk score
Target condition and reference standard(s)	Difficult laryngoscopy: Cormack and Lehane
Flow and timing	Index tests performed 2 days before general surgery
Comparative	
Notes	
Methodological quality	

Yamamoto 1997 (*Continued*)

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?			
Was a case-control design avoided?			
Did the study avoid inappropriate exclusions?			
DOMAIN 2: Index Test All Tests			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Did the assessors of the index test have appropriate training?			
Was interobserver variability reported for some or all patients?			
Was interobserver agreement acceptable?			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?			
Were the reference standard results interpreted without knowledge of the results of the index tests?			

Yamamoto 1997 (*Continued*)

DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?			
Did all patients receive the same reference standard?			
Were all patients included in the analysis?			

Yildiz 2005

Study characteristics	
Patient sampling	Unknown
Patient characteristics and setting	Patients presenting to a Turkish hospital. No other details Sample size: 576 346 females Mean age: 45 years
Index tests	MMT, combination of tests
Target condition and reference standard(s)	Difficult face mask ventilation
Flow and timing	Unclear
Comparative	
Notes	

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?			
Was a case-control design avoided?			

Did the study avoid inappropriate exclusions?			
DOMAIN 2: Index Test All Tests			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Did the assessors of the index test have appropriate training?			
Was interobserver variability reported for some or all patients?			
Was interobserver agreement acceptable?			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?			
Were the reference standard results interpreted without knowledge of the results of the index tests?			
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?			
Did all patients receive the same reference standard?			

Were all patients included in the analysis?			

Yildiz 2007

Study characteristics			
Patient sampling	ASA I-III patients scheduled for elective surgery and requiring endotracheal intubation from seven sites. Patients aged > 18 years, those requiring RSI or an awake intubation, obstetrical patients, surgical procedures involving the upper airway, or patients with a history of difficult intubation or tracheotomy were excluded from the study. No informaiton on selection		
Patient characteristics and setting	Sample size: 1700 994 females Mean age: 43.5 years		
Index tests	MMT, TMD (< 4.8 cm), SMD (< 10.5 cm), mouth opening (< 4 cm), combination of tests		
Target condition and reference standard(s)	Difficult laryngoscopy: Cormack and Lehane; difficult face mask ventilation		
Flow and timing	Index tests: preoperative visit Reference standard: after induction of GA		
Comparative			
Notes			
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?			
Was a case-control design avoided?			
Did the study avoid inappropriate exclusions?			
DOMAIN 2: Index Test All Tests			

Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			
Did the assessors of the index test have appropriate training?			
Was interobserver variability reported for some or all patients?			
Was interobserver agreement acceptable?			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?			
Were the reference standard results interpreted without knowledge of the results of the index tests?			
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?			
Did all patients receive the same reference standard?			
Were all patients included in the analysis?			

Yu 2015

Study characteristics			
Patient sampling	"This prospective, observational study was conducted among patients who had been admitted to our 20-bed operation center in a university hospital During the observation, 1200 patients scheduled to undergo general anesthesia with endotracheal intubation for elective surgery were screened."		
Patient characteristics and setting	All Chinese patients Sample Size: 732 358 females Mean age 50.8 years		
Index tests	Combination of tests		
Target condition and reference standard(s)	Difficult laryngoscopy: Cormack & Lehane; Difficult tracheal intubation		
Flow and timing			
Comparative			
Notes			
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?			
Was a case-control design avoided?			
Did the study avoid inappropriate exclusions?			
DOMAIN 2: Index Test All Tests			
Were the index test results interpreted without knowledge of the results of the reference standard?			
If a threshold was used, was it pre-specified?			

Did the assessors of the index test have appropriate training?			
Was interobserver variability reported for some or all patients?			
Was interobserver agreement acceptable?			
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?			
Were the reference standard results interpreted without knowledge of the results of the index tests?			
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?			
Did all patients receive the same reference standard?			
Were all patients included in the analysis?			

ASA: American Society of Anesthesiologists Physical Status; BMI: body mass index; BURP: backward, upward and rightward pressure; DTA: diagnostic test accuracy; ED: emergency department; EMS: emergency medical services; ENT: ear, nose and throat; ETI: endotracheal intubation; GA: general anaesthesia; ICU: intensive care unit; IDS: intubation difficulty scale; MT: Mallampati test; MMT: modified Mallampati test; OP: operation; OR: operating room; RSI: rapid sequence induction; SMD: sternomental distance; TMD: thyromental distance; ULBT: upper lip bite test

Characteristics of excluded studies *[ordered by study ID]*

Study	Reason for exclusion
Acer 2011	Insufficient data to calculate measures of diagnostic test accuracy
Acikgoz 2015	Insufficient data to calculate measures of diagnostic test accuracy
Beyus 2010	Insufficient data to calculate measures of diagnostic test accuracy
Hiremath 1998	Insufficient data to calculate measures of diagnostic test accuracy
Lewis 1994	Insufficient data to calculate measures of diagnostic test accuracy
Meininger 2010	Insufficient data to calculate measures of diagnostic test accuracy
Moon 2013	Insufficient data to calculate measures of diagnostic test accuracy
Oriol-Lo pez 2009	Insufficient data to calculate measures of diagnostic test accuracy
Orozco-Di az 2010	Insufficient data to calculate measures of diagnostic test accuracy
Reed 2005	Insufficient data to calculate measures of diagnostic test accuracy
Safavi 2011	Insufficient data to calculate measures of diagnostic test accuracy
Siyam 2002	Insufficient data to calculate measures of diagnostic test accuracy
Tripathi 2006	Insufficient data to calculate measures of diagnostic test accuracy

Characteristics of studies awaiting classification *[ordered by study ID]*

Akhlaghi 2017

Study characteristics	
Patient sampling	
Patient characteristics and setting	
Index tests	
Target condition and reference standard(s)	
Flow and timing	

Akhlaghi 2017 (*Continued*)

Comparative	
Notes	Result from top-up search; will be incorporated into the review at the next update

Andrade 2017

Study characteristics	
Patient sampling	
Patient characteristics and setting	
Index tests	
Target condition and reference standard(s)	
Flow and timing	
Comparative	
Notes	Result from top-up search; will be incorporated into the review at the next update

Awan 2017

Study characteristics	
Patient sampling	
Patient characteristics and setting	
Index tests	
Target condition and reference standard(s)	
Flow and timing	
Comparative	
Notes	Result from top-up search; will be incorporated into the review at the next update

Banik 2017

Study characteristics	
Patient sampling	
Patient characteristics and setting	
Index tests	
Target condition and reference standard(s)	
Flow and timing	
Comparative	
Notes	Result from top-up search; will be incorporated into the review at the next update

Belda 2017

Study characteristics	
Patient sampling	
Patient characteristics and setting	
Index tests	
Target condition and reference standard(s)	
Flow and timing	
Comparative	
Notes	Result from top-up search; will be incorporated into the review at the next update

Card 2017

Study characteristics	
Patient sampling	
Patient characteristics and setting	

Card 2017 (*Continued*)

Index tests	
Target condition and reference standard(s)	
Flow and timing	
Comparative	
Notes	Result from top-up search; will be incorporated into the review at the next update

Carlson 2017

Study characteristics	
Patient sampling	
Patient characteristics and setting	
Index tests	
Target condition and reference standard(s)	
Flow and timing	
Comparative	
Notes	Result from top-up search; will be incorporated into the review at the next update

Dar 2017

Study characteristics	
Patient sampling	
Patient characteristics and setting	
Index tests	
Target condition and reference standard(s)	
Flow and timing	

Dar 2017 (*Continued*)

Comparative	
Notes	Result from top-up search; will be incorporated into the review at the next update

Eiamcharoenwit 2017

Study characteristics	
Patient sampling	
Patient characteristics and setting	
Index tests	
Target condition and reference standard(s)	
Flow and timing	
Comparative	
Notes	Result from top-up search; will be incorporated into the review at the next update

Han 2017

Study characteristics	
Patient sampling	
Patient characteristics and setting	
Index tests	
Target condition and reference standard(s)	
Flow and timing	
Comparative	
Notes	Result from top-up search; will be incorporated into the review at the next update

Hanouz 2018

Study characteristics	
Patient sampling	
Patient characteristics and setting	
Index tests	
Target condition and reference standard(s)	
Flow and timing	
Comparative	
Notes	Result from top-up search; will be incorporated into the review at the next update

Jain 2017

Study characteristics	
Patient sampling	
Patient characteristics and setting	
Index tests	
Target condition and reference standard(s)	
Flow and timing	
Comparative	
Notes	Result from top-up search; will be incorporated into the review at the next update

Khatiwada 2017

Study characteristics	
Patient sampling	
Patient characteristics and setting	

Khatiwada 2017 (*Continued*)

Index tests	
Target condition and reference standard(s)	
Flow and timing	
Comparative	
Notes	Result from top-up search; will be incorporated into the review at the next update

Lee 2017

Study characteristics	
Patient sampling	
Patient characteristics and setting	
Index tests	
Target condition and reference standard(s)	
Flow and timing	
Comparative	
Notes	Result from top-up search; will be incorporated into the review at the next update

Mahmoodpoor 2017

Study characteristics	
Patient sampling	
Patient characteristics and setting	
Index tests	
Target condition and reference standard(s)	
Flow and timing	

Mahmoodpoor 2017 (*Continued*)

Comparative	
Notes	Result from top-up search; will be incorporated into the review at the next update

Norskov 2017

Study characteristics	
Patient sampling	
Patient characteristics and setting	
Index tests	
Target condition and reference standard(s)	
Flow and timing	
Comparative	
Notes	Result from top-up search; will be incorporated into the review at the next update

Prakash 2017

Study characteristics	
Patient sampling	
Patient characteristics and setting	
Index tests	
Target condition and reference standard(s)	
Flow and timing	
Comparative	
Notes	Result from top-up search; will be incorporated into the review at the next update

Rao 2017

Study characteristics	
Patient sampling	
Patient characteristics and setting	
Index tests	
Target condition and reference standard(s)	
Flow and timing	
Comparative	
Notes	Result from top-up search; will be incorporated into the review at the next update

Riad 2018

Study characteristics	
Patient sampling	
Patient characteristics and setting	
Index tests	
Target condition and reference standard(s)	
Flow and timing	
Comparative	
Notes	Result from top-up search; will be incorporated into the review at the next update

Selvi 2017

Study characteristics	
Patient sampling	
Patient characteristics and setting	

Selvi 2017 (*Continued*)

Index tests	
Target condition and reference standard(s)	
Flow and timing	
Comparative	
Notes	Result from top-up search; will be incorporated into the review at the next update

Shankar 2017

Study characteristics	
Patient sampling	
Patient characteristics and setting	
Index tests	
Target condition and reference standard(s)	
Flow and timing	
Comparative	
Notes	Result from top-up search; will be incorporated into the review at the next update

Siljeblad 2017

Study characteristics	
Patient sampling	
Patient characteristics and setting	
Index tests	
Target condition and reference standard(s)	
Flow and timing	

Siljeblad 2017 (*Continued*)

Comparative	
Notes	Result from top-up search; will be incorporated into the review at the next update

Srivilaithon 2018

Study characteristics	
Patient sampling	
Patient characteristics and setting	
Index tests	
Target condition and reference standard(s)	
Flow and timing	
Comparative	
Notes	Result from top-up search; will be incorporated into the review at the next update

Torres 2017

Study characteristics	
Patient sampling	
Patient characteristics and setting	
Index tests	
Target condition and reference standard(s)	
Flow and timing	
Comparative	
Notes	Result from top-up search; will be incorporated into the review at the next update

Wang 2017

Study characteristics	
Patient sampling	
Patient characteristics and setting	
Index tests	
Target condition and reference standard(s)	
Flow and timing	
Comparative	
Notes	Result from top-up search; will be incorporated into the review at the next update

Workeneh 2017

Study characteristics	
Patient sampling	
Patient characteristics and setting	
Index tests	
Target condition and reference standard(s)	
Flow and timing	
Comparative	
Notes	Result from top-up search; will be incorporated into the review at the next update

Yildirim 2017

Study characteristics	
Patient sampling	
Patient characteristics and setting	

Yildirim 2017 (*Continued*)

Index tests	
Target condition and reference standard(s)	
Flow and timing	
Comparative	
Notes	Result from top-up search; will be incorporated into the review at the next update

DATA

Presented below are all the data for all of the tests entered into the review.

Tests. Data tables by test

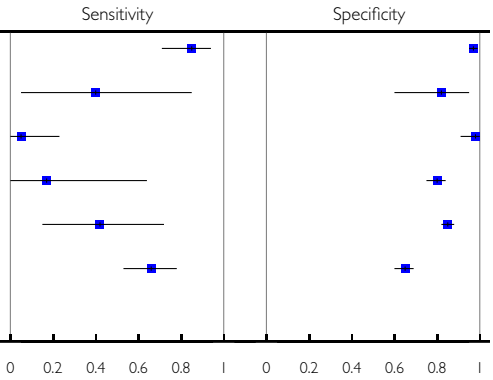
Test	No. of studies	No. of participants
1 Mallampati test: difficult laryngoscopy	6	2165
2 Mallampati test: difficult tracheal intubation	1	500
3 Modified Mallampati test: difficult laryngoscopy	80	232939
4 Modified Mallampati test: difficult face mask ventilation	6	56323
5 Modified Mallampati test: difficult tracheal intubation	24	191849
6 Modified Mallampati test: failed intubation	2	485
7 Wilson risk score: difficult laryngoscopy	5	5862
8 Wilson risk score: difficult tracheal intubation	1	123
9 Thyromental distance: difficult laryngoscopy	42	33189
10 Thyromental distance: difficult face mask ventilation	1	53041
11 Thyromental distance: difficult tracheal intubation	10	5089
12 Sternomental distance: difficult laryngoscopy	16	12211
13 Sternomental distance: difficult tracheal intubation	2	864
14 Mouth opening: difficult laryngoscopy	24	22179
15 Mouth opening: difficult face mask ventilation	2	53469
16 Mouth opening: difficult tracheal intubation	9	6091
17 Upper lip bite test: difficult laryngoscopy	27	19609
18 Upper lip bite test: difficult face mask ventilation	1	200
19 Upper lip bite test: difficult tracheal intubation	2	598
20 Combination of tests: difficult laryngoscopy	42	230680

Test 1. Mallampati test: difficult laryngoscopy.

Review: Airway physical examination tests for detection of difficult airway management in apparently normal adult patients

Test: 1 Mallampati test: difficult laryngoscopy

Study	TP	FP	FN	TN	Sensitivity	Specificity
Bilgin 1998	39	12	7	442	0.85 [0.71, 0.94]	0.97 [0.95, 0.99]
Hui 2009	2	4	3	18	0.40 [0.05, 0.85]	0.82 [0.60, 0.95]
Nadal 1998	1	1	21	60	0.05 [0.00, 0.23]	0.98 [0.91, 1.00]
Oates 1990	1	65	5	262	0.17 [0.00, 0.64]	0.80 [0.75, 0.84]
Oates 1991	5	109	7	630	0.42 [0.15, 0.72]	0.85 [0.82, 0.88]
Tse 1995	41	145	21	264	0.66 [0.53, 0.78]	0.65 [0.60, 0.69]

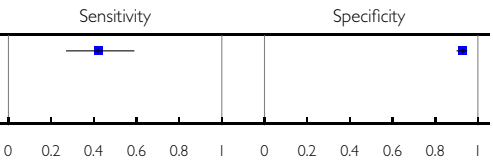


Test 2. Mallampati test: difficult tracheal intubation.

Review: Airway physical examination tests for detection of difficult airway management in apparently normal adult patients

Test: 2 Mallampati test: difficult tracheal intubation

Study	TP	FP	FN	TN	Sensitivity	Specificity
Bilgin 1998	17	34	23	426	0.43 [0.27, 0.59]	0.93 [0.90, 0.95]

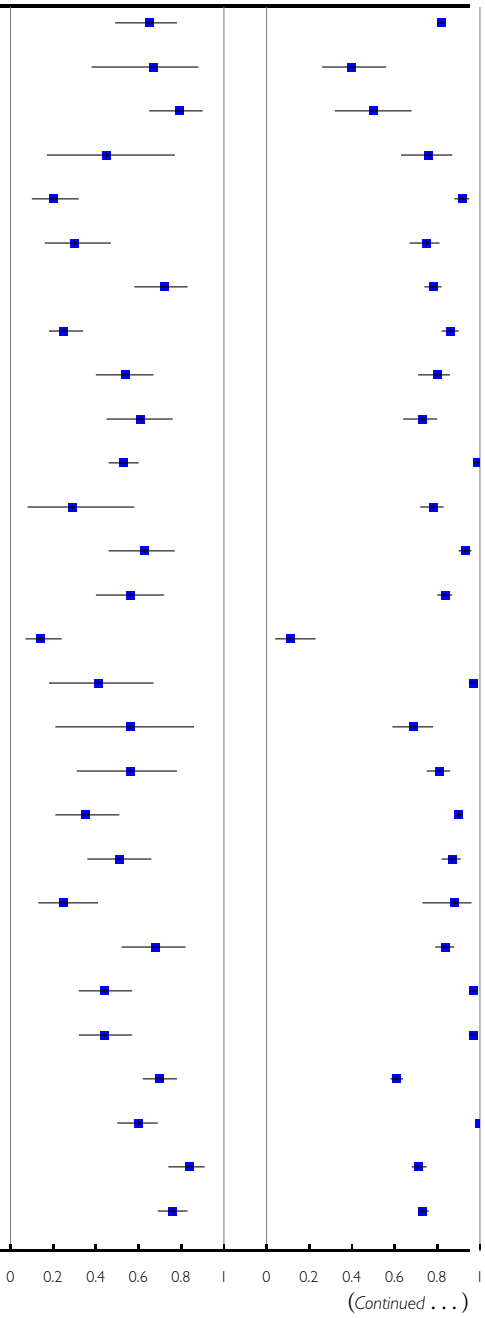


Test 3. Modified Mallampati test: difficult laryngoscopy.

Review: Airway physical examination tests for detection of difficult airway management in apparently normal adult patients

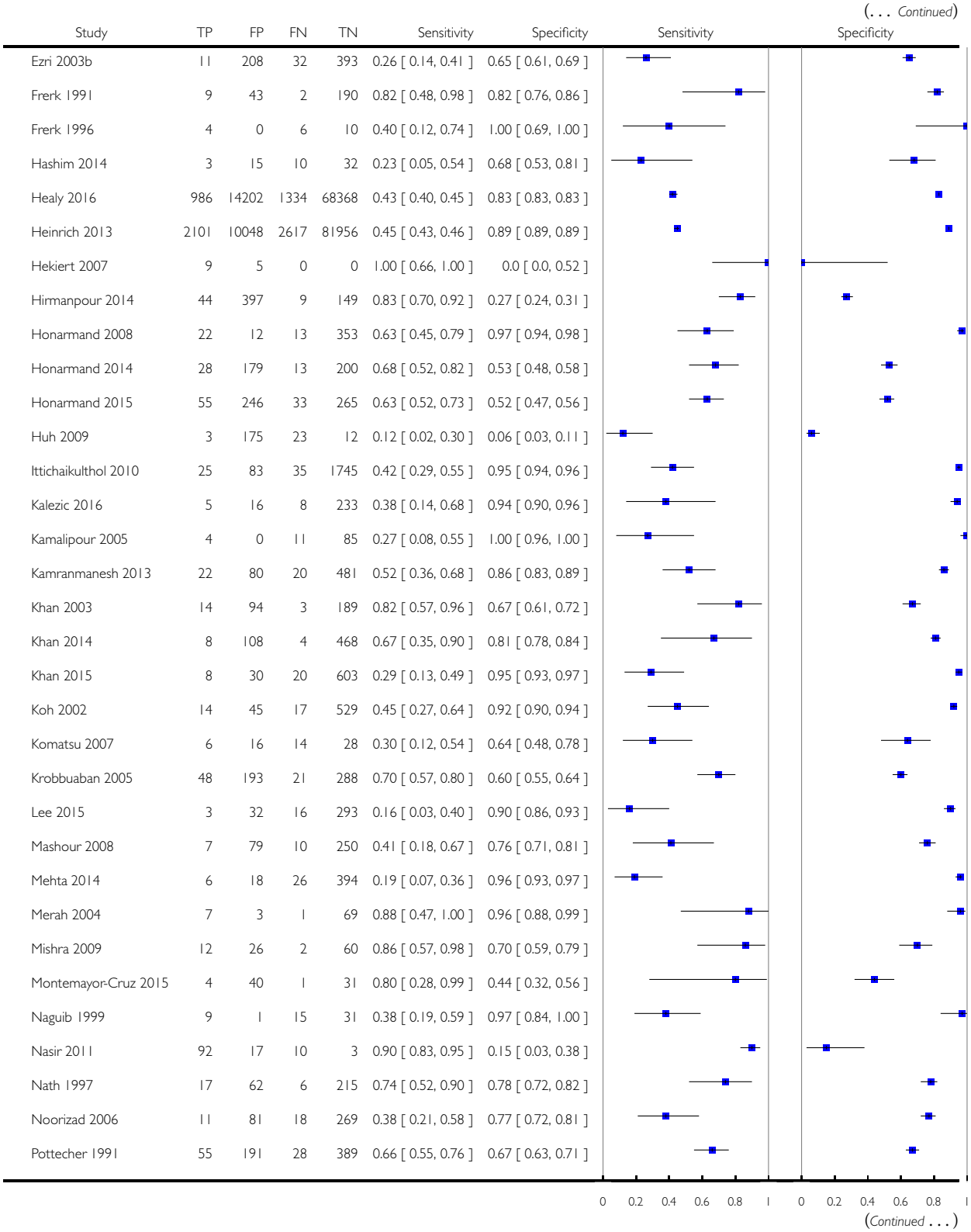
Test: 3 Modified Mallampati test: difficult laryngoscopy

Study	TP	FP	FN	TN	Sensitivity	Specificity
Adamus 2010	31	258	17	1212	0.65 [0.49, 0.78]	0.82 [0.80, 0.84]
Adnet 2001	10	28	5	19	0.67 [0.38, 0.88]	0.40 [0.26, 0.56]
Aktas 2015	38	17	10	17	0.79 [0.65, 0.90]	0.50 [0.32, 0.68]
Ali 2009	5	13	6	42	0.45 [0.17, 0.77]	0.76 [0.63, 0.87]
Ali 2012	11	22	45	246	0.20 [0.10, 0.32]	0.92 [0.88, 0.95]
Allahyary 2008	11	42	26	124	0.30 [0.16, 0.47]	0.75 [0.67, 0.81]
Ambesh 2013	38	97	15	350	0.72 [0.58, 0.83]	0.78 [0.74, 0.82]
Ayhan 2016	31	42	92	262	0.25 [0.18, 0.34]	0.86 [0.82, 0.90]
Ayuso 2009	29	26	25	101	0.54 [0.40, 0.67]	0.80 [0.71, 0.86]
Badheka 2016	25	35	16	94	0.61 [0.45, 0.76]	0.73 [0.64, 0.80]
Baig 2014	99	2	88	230	0.53 [0.46, 0.60]	0.99 [0.97, 1.00]
Basaranoglu 2010	4	49	10	176	0.29 [0.08, 0.58]	0.78 [0.72, 0.83]
Basunia 2013	25	17	15	243	0.63 [0.46, 0.77]	0.93 [0.90, 0.96]
Bhat 2007	22	76	17	385	0.56 [0.40, 0.72]	0.84 [0.80, 0.87]
Bindra 2010	11	42	65	5	0.14 [0.07, 0.24]	0.11 [0.04, 0.23]
Bouaggad 2004	7	8	10	295	0.41 [0.18, 0.67]	0.97 [0.95, 0.99]
Brodsky 2002	5	28	4	63	0.56 [0.21, 0.86]	0.69 [0.59, 0.78]
Butler 1992	10	38	8	164	0.56 [0.31, 0.78]	0.81 [0.75, 0.86]
Cattano 2004	15	182	28	1731	0.35 [0.21, 0.51]	0.90 [0.89, 0.92]
Choi 2013	25	29	24	191	0.51 [0.36, 0.66]	0.87 [0.82, 0.91]
Connor 2011	10	5	30	35	0.25 [0.13, 0.41]	0.88 [0.73, 0.96]
Descoins 1994	28	41	13	213	0.68 [0.52, 0.82]	0.84 [0.79, 0.88]
Domi 2009	30	10	38	348	0.44 [0.32, 0.57]	0.97 [0.95, 0.99]
Domi 2010	30	10	38	348	0.44 [0.32, 0.57]	0.97 [0.95, 0.99]
Eberhart 2005	92	381	39	595	0.70 [0.62, 0.78]	0.61 [0.58, 0.64]
El-Ganzouri 1996	64	13	43	9097	0.60 [0.50, 0.69]	1.00 [1.00, 1.00]
Ezri 2001	68	196	13	487	0.84 [0.74, 0.91]	0.71 [0.68, 0.75]
Ezri 2003a	116	354	36	966	0.76 [0.69, 0.83]	0.73 [0.71, 0.76]



(*Continued . . .*)

(. . . *Continued*)

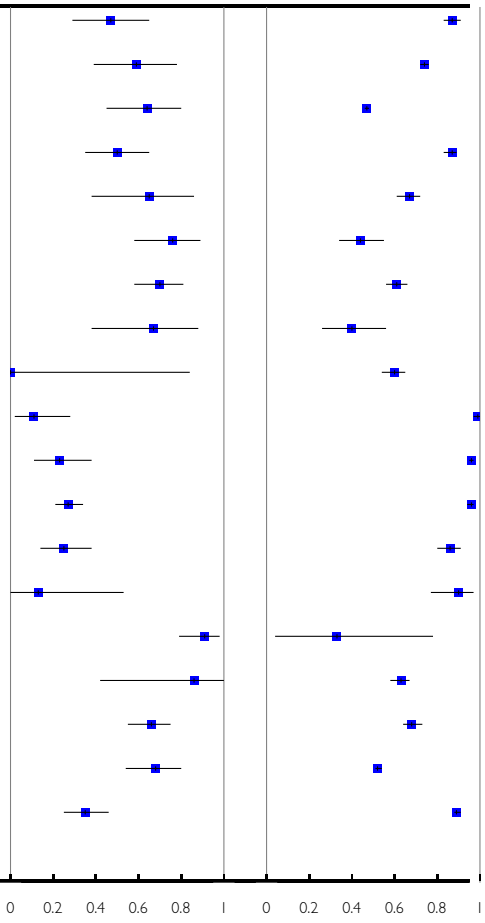


Study	TP	FP	FN	TN	Sensitivity	Specificity
Ezri 2003b	11	208	32	393	0.26 [0.14, 0.41]	0.65 [0.61, 0.69]
Frerk 1991	9	43	2	190	0.82 [0.48, 0.98]	0.82 [0.76, 0.86]
Frerk 1996	4	0	6	10	0.40 [0.12, 0.74]	1.00 [0.69, 1.00]
Hashim 2014	3	15	10	32	0.23 [0.05, 0.54]	0.68 [0.53, 0.81]
Healy 2016	986	14202	1334	68368	0.43 [0.40, 0.45]	0.83 [0.83, 0.83]
Heinrich 2013	2101	10048	2617	81956	0.45 [0.43, 0.46]	0.89 [0.89, 0.89]
Hekiert 2007	9	5	0	0	1.00 [0.66, 1.00]	0.0 [0.0, 0.52]
Hirmanpour 2014	44	397	9	149	0.83 [0.70, 0.92]	0.27 [0.24, 0.31]
Honarmand 2008	22	12	13	353	0.63 [0.45, 0.79]	0.97 [0.94, 0.98]
Honarmand 2014	28	179	13	200	0.68 [0.52, 0.82]	0.53 [0.48, 0.58]
Honarmand 2015	55	246	33	265	0.63 [0.52, 0.73]	0.52 [0.47, 0.56]
Huh 2009	3	175	23	12	0.12 [0.02, 0.30]	0.06 [0.03, 0.11]
Ittichaikulthol 2010	25	83	35	1745	0.42 [0.29, 0.55]	0.95 [0.94, 0.96]
Kalezic 2016	5	16	8	233	0.38 [0.14, 0.68]	0.94 [0.90, 0.96]
Kamalipour 2005	4	0	11	85	0.27 [0.08, 0.55]	1.00 [0.96, 1.00]
Kamranmanesh 2013	22	80	20	481	0.52 [0.36, 0.68]	0.86 [0.83, 0.89]
Khan 2003	14	94	3	189	0.82 [0.57, 0.96]	0.67 [0.61, 0.72]
Khan 2014	8	108	4	468	0.67 [0.35, 0.90]	0.81 [0.78, 0.84]
Khan 2015	8	30	20	603	0.29 [0.13, 0.49]	0.95 [0.93, 0.97]
Koh 2002	14	45	17	529	0.45 [0.27, 0.64]	0.92 [0.90, 0.94]
Komatsu 2007	6	16	14	28	0.30 [0.12, 0.54]	0.64 [0.48, 0.78]
Krobbuaban 2005	48	193	21	288	0.70 [0.57, 0.80]	0.60 [0.55, 0.64]
Lee 2015	3	32	16	293	0.16 [0.03, 0.40]	0.90 [0.86, 0.93]
Mashour 2008	7	79	10	250	0.41 [0.18, 0.67]	0.76 [0.71, 0.81]
Mehta 2014	6	18	26	394	0.19 [0.07, 0.36]	0.96 [0.93, 0.97]
Merah 2004	7	3	1	69	0.88 [0.47, 1.00]	0.96 [0.88, 0.99]
Mishra 2009	12	26	2	60	0.86 [0.57, 0.98]	0.70 [0.59, 0.79]
Montemayor-Cruz 2015	4	40	1	31	0.80 [0.28, 0.99]	0.44 [0.32, 0.56]
Naguib 1999	9	1	15	31	0.38 [0.19, 0.59]	0.97 [0.84, 1.00]
Nasir 2011	92	17	10	3	0.90 [0.83, 0.95]	0.15 [0.03, 0.38]
Nath 1997	17	62	6	215	0.74 [0.52, 0.90]	0.78 [0.72, 0.82]
Noorizad 2006	11	81	18	269	0.38 [0.21, 0.58]	0.77 [0.72, 0.81]
Pottecher 1991	55	191	28	389	0.66 [0.55, 0.76]	0.67 [0.63, 0.71]

(*Continued* . . .)

(. . . *Continued*)

Study	TP	FP	FN	TN	Sensitivity	Specificity
Prakash 2013	15	38	17	260	0.47 [0.29, 0.65]	0.87 [0.83, 0.91]
Rocke 1992	16	381	11	1092	0.59 [0.39, 0.78]	0.74 [0.72, 0.76]
Safavi 2014	21	2540	12	2248	0.64 [0.45, 0.80]	0.47 [0.46, 0.48]
Samra 1995	24	69	24	449	0.50 [0.35, 0.65]	0.87 [0.83, 0.89]
Savva 1994	11	111	6	222	0.65 [0.38, 0.86]	0.67 [0.61, 0.72]
Schmitt 2000	25	53	8	42	0.76 [0.58, 0.89]	0.44 [0.34, 0.55]
Shah 2013	47	161	20	252	0.70 [0.58, 0.81]	0.61 [0.56, 0.66]
Sharma 2010	10	28	5	19	0.67 [0.38, 0.88]	0.40 [0.26, 0.56]
Singh 2009	0	120	2	178	0.0 [0.0, 0.84]	0.60 [0.54, 0.65]
Tantri 2016	3	2	25	247	0.11 [0.02, 0.28]	0.99 [0.97, 1.00]
Thompson 2009	9	67	31	1504	0.23 [0.11, 0.38]	0.96 [0.95, 0.97]
Ul Haq 2013	46	25	124	565	0.27 [0.21, 0.34]	0.96 [0.94, 0.97]
Vallem 2015	13	20	40	127	0.25 [0.14, 0.38]	0.86 [0.80, 0.91]
Vani 2000	1	4	7	38	0.13 [0.00, 0.53]	0.90 [0.77, 0.97]
Wajekar 2015	42	4	4	2	0.91 [0.79, 0.98]	0.33 [0.04, 0.78]
Wong 1999	6	151	1	253	0.86 [0.42, 1.00]	0.63 [0.58, 0.67]
Wong 2009	57	151	30	327	0.66 [0.55, 0.75]	0.68 [0.64, 0.73]
Yamamoto 1997	38	1723	18	1901	0.68 [0.54, 0.80]	0.52 [0.51, 0.54]
Yildiz 2007	28	172	52	1422	0.35 [0.25, 0.46]	0.89 [0.88, 0.91]

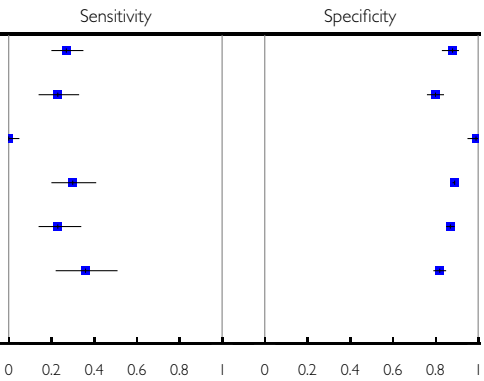


Test 4. Modified Mallampati test: difficult face mask ventilation.

Review: Airway physical examination tests for detection of difficult airway management in apparently normal adult patients

Test: 4 Modified Mallampati test: difficult face mask ventilation

Study	TP	FP	FN	TN	Sensitivity	Specificity
Ayhan 2016	38	35	102	252	0.27 [0.20, 0.35]	0.88 [0.83, 0.91]
Cattano 2014	18	95	62	382	0.23 [0.14, 0.33]	0.80 [0.76, 0.84]
Khan 2009b	0	2	76	142	0.0 [0.0, 0.05]	0.99 [0.95, 1.00]
Kheterpal 2009	23	5950	54	47014	0.30 [0.20, 0.41]	0.89 [0.88, 0.89]
Langeron 2000	17	186	58	1241	0.23 [0.14, 0.34]	0.87 [0.85, 0.89]
Yildiz 2005	16	94	29	437	0.36 [0.22, 0.51]	0.82 [0.79, 0.85]

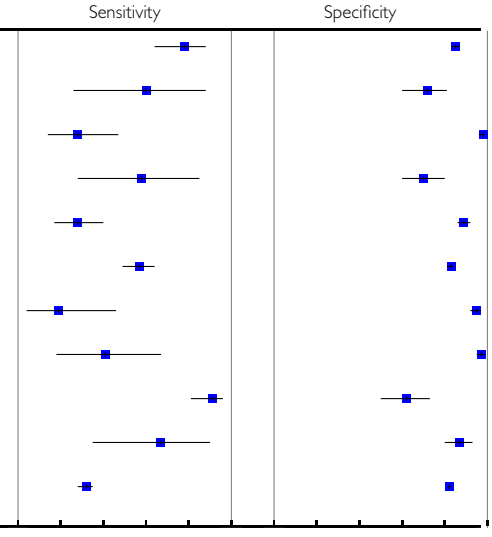


Test 5. Modified Mallampati test: difficult tracheal intubation.

Review: Airway physical examination tests for detection of difficult airway management in apparently normal adult patients

Test: 5 Modified Mallampati test: difficult tracheal intubation

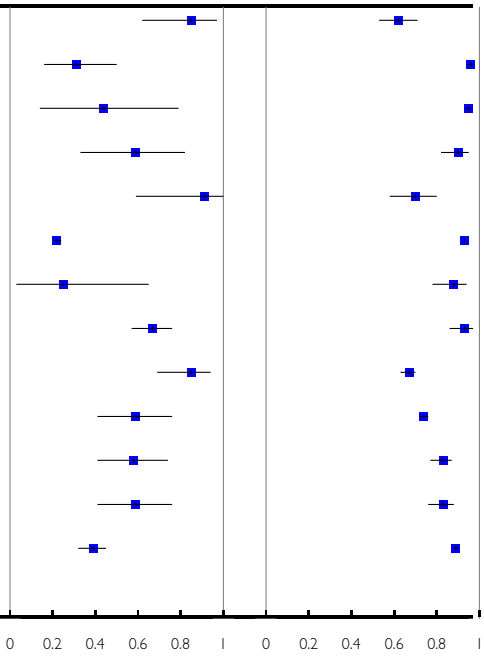
Study	TP	FP	FN	TN	Sensitivity	Specificity
Arne 1998	39	168	11	982	0.78 [0.64, 0.88]	0.85 [0.83, 0.87]
Bergler 1997	6	23	4	58	0.60 [0.26, 0.88]	0.72 [0.60, 0.81]
Bouaggad 2004	9	6	23	282	0.28 [0.14, 0.47]	0.98 [0.96, 0.99]
Brodsky 2002	7	26	5	62	0.58 [0.28, 0.85]	0.70 [0.60, 0.80]
Chaves 2009	18	49	47	398	0.28 [0.17, 0.40]	0.89 [0.86, 0.92]
De Jong 2015	111	346	85	1673	0.57 [0.49, 0.64]	0.83 [0.81, 0.84]
Dohrn 2015	3	24	13	431	0.19 [0.04, 0.46]	0.95 [0.92, 0.97]
Ezri 2003c	7	8	10	295	0.41 [0.18, 0.67]	0.97 [0.95, 0.99]
Fritscherova 2011	67	28	7	46	0.91 [0.81, 0.96]	0.62 [0.50, 0.73]
Gonzalez 2008	8	15	4	104	0.67 [0.35, 0.90]	0.87 [0.80, 0.93]
Healy 2016	287	14901	624	69764	0.32 [0.28, 0.35]	0.82 [0.82, 0.83]



(*Continued* . . .)

(. . . *Continued*)

Study	TP	FP	FN	TN	Sensitivity	Specificity
Juvin 2003	17	41	3	68	0.85 [0.62, 0.97]	0.62 [0.53, 0.71]
K Nasa 2014	10	22	22	530	0.31 [0.16, 0.50]	0.96 [0.94, 0.97]
Khan 2015	4	34	5	618	0.44 [0.14, 0.79]	0.95 [0.93, 0.96]
Kim 2011	10	11	7	95	0.59 [0.33, 0.82]	0.90 [0.82, 0.95]
Knudsen 2014	10	23	1	53	0.91 [0.59, 1.00]	0.70 [0.58, 0.80]
Lundstrom 2009	1042	5898	3620	81531	0.22 [0.21, 0.24]	0.93 [0.93, 0.93]
Mallat 2010	2	9	6	63	0.25 [0.03, 0.65]	0.88 [0.78, 0.94]
Naguib 2006	65	7	32	90	0.67 [0.57, 0.76]	0.93 [0.86, 0.97]
Pottecher 1991	33	209	6	415	0.85 [0.69, 0.94]	0.67 [0.63, 0.70]
Rocke 1992	19	378	13	1090	0.59 [0.41, 0.76]	0.74 [0.72, 0.76]
Seo 2012	21	47	15	222	0.58 [0.41, 0.74]	0.83 [0.77, 0.87]
Tuzuner-Oncul 2008	19	30	13	144	0.59 [0.41, 0.76]	0.83 [0.76, 0.88]
Uribe 2015	87	435	138	3482	0.39 [0.32, 0.45]	0.89 [0.88, 0.90]

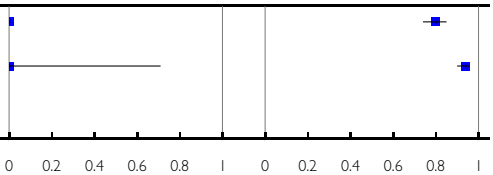


Test 6. Modified Mallampati test: failed intubation.

Review: Airway physical examination tests for detection of difficult airway management in apparently normal adult patients

Test: 6 Modified Mallampati test: failed intubation

Study	TP	FP	FN	TN	Sensitivity	Specificity
Hagberg 2009	0	50	0	196	0.0 [0.0, 0.0]	0.80 [0.74, 0.85]
Istvan 2010	0	15	3	221	0.0 [0.0, 0.71]	0.94 [0.90, 0.96]

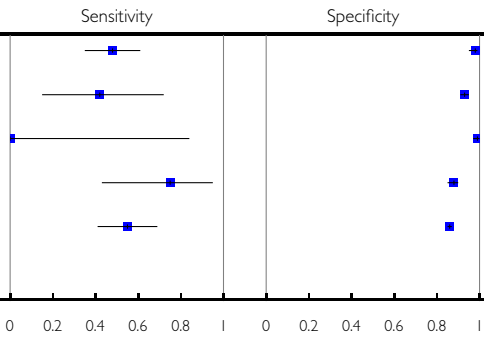


Test 7. Wilson risk score: difficult laryngoscopy.

Review: Airway physical examination tests for detection of difficult airway management in apparently normal adult patients

Test: 7 Wilson risk score: difficult laryngoscopy

Study	TP	FP	FN	TN	Sensitivity	Specificity
Domi 2009	30	9	33	353	0.48 [0.35, 0.61]	0.98 [0.95, 0.99]
Oates 1991	5	51	7	688	0.42 [0.15, 0.72]	0.93 [0.91, 0.95]
Singh 2009	0	3	2	295	0.0 [0.0, 0.84]	0.99 [0.97, 1.00]
Wilson 1988	9	93	3	673	0.75 [0.43, 0.95]	0.88 [0.85, 0.90]
Yamamoto 1997	31	493	25	3059	0.55 [0.41, 0.69]	0.86 [0.85, 0.87]

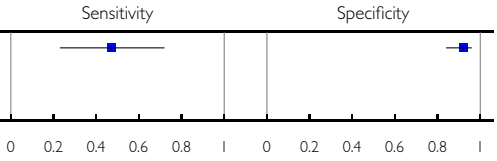


Test 8. Wilson risk score: difficult tracheal intubation.

Review: Airway physical examination tests for detection of difficult airway management in apparently normal adult patients

Test: 8 Wilson risk score: difficult tracheal intubation

Study	TP	FP	FN	TN	Sensitivity	Specificity
Kim 2011	8	9	9	97	0.47 [0.23, 0.72]	0.92 [0.84, 0.96]

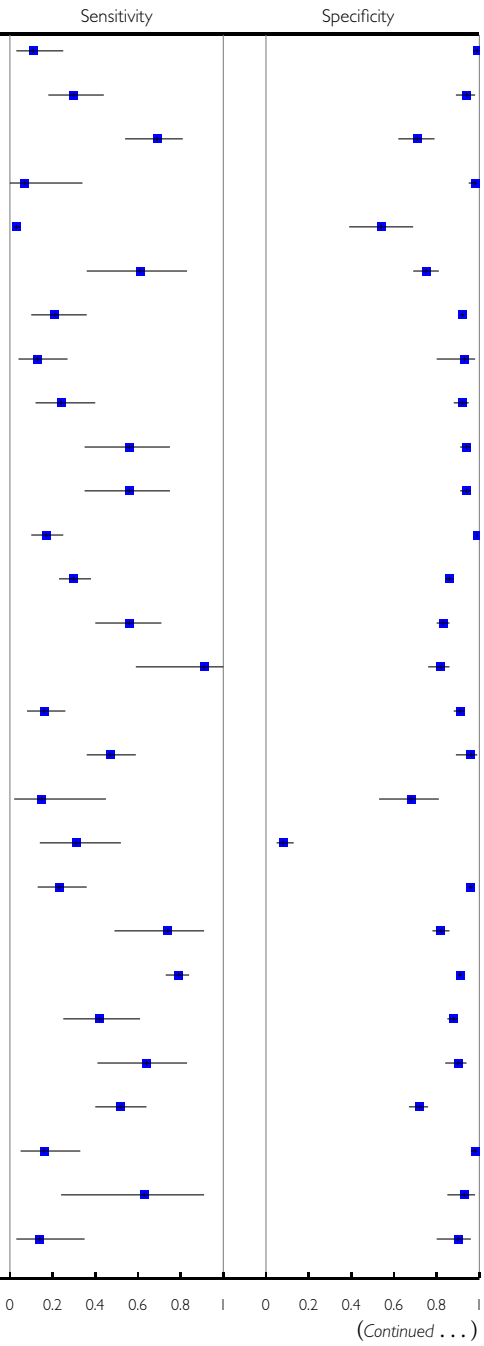


Test 9. Thyromental distance: difficult laryngoscopy.

Review: Airway physical examination tests for detection of difficult airway management in apparently normal adult patients

Test: 9 Thyromental distance: difficult laryngoscopy

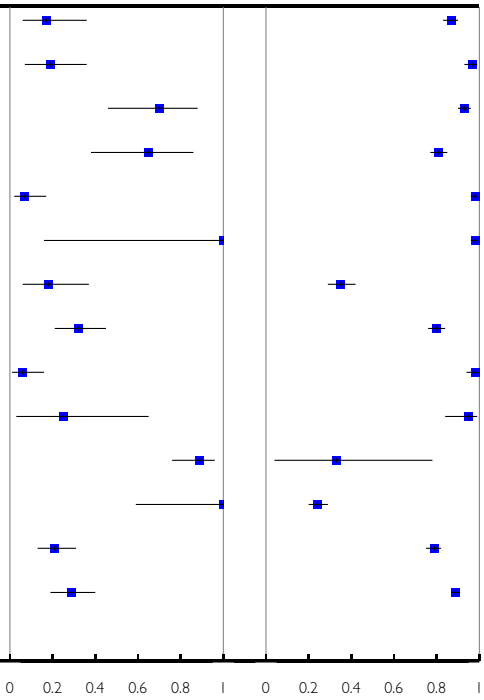
Study	TP	FP	FN	TN	Sensitivity	Specificity
Allahyary 2008	4	1	33	165	0.11 [0.03, 0.25]	0.99 [0.97, 1.00]
Ayuso 2009	16	7	38	120	0.30 [0.18, 0.44]	0.94 [0.89, 0.98]
Badheka 2016	35	35	16	84	0.69 [0.54, 0.81]	0.71 [0.62, 0.79]
Basaranoglu 2010	1	5	13	220	0.07 [0.00, 0.34]	0.98 [0.95, 0.99]
Bilgin 1998	14	22	438	26	0.03 [0.02, 0.05]	0.54 [0.39, 0.69]
Butler 1992	11	50	7	152	0.61 [0.36, 0.83]	0.75 [0.69, 0.81]
Cattano 2004	9	153	34	1760	0.21 [0.10, 0.36]	0.92 [0.91, 0.93]
Connor 2011	5	3	35	37	0.13 [0.04, 0.27]	0.93 [0.80, 0.98]
Descoins 1994	10	20	31	234	0.24 [0.12, 0.40]	0.92 [0.88, 0.95]
Domi 2009	15	25	12	374	0.56 [0.35, 0.75]	0.94 [0.91, 0.96]
Domi 2010	15	25	12	374	0.56 [0.35, 0.75]	0.94 [0.91, 0.96]
El-Ganzouri 1996	18	99	89	10301	0.17 [0.10, 0.25]	0.99 [0.99, 0.99]
Ezri 2003a	46	184	106	1136	0.30 [0.23, 0.38]	0.86 [0.84, 0.88]
Ezri 2003b	24	101	19	500	0.56 [0.40, 0.71]	0.83 [0.80, 0.86]
Frerk 1991	10	43	1	190	0.91 [0.59, 1.00]	0.82 [0.76, 0.86]
Freund 2012	11	52	59	531	0.16 [0.08, 0.26]	0.91 [0.88, 0.93]
Fritscherova 2011	35	3	39	71	0.47 [0.36, 0.59]	0.96 [0.89, 0.99]
Hashim 2014	2	15	11	32	0.15 [0.02, 0.45]	0.68 [0.53, 0.81]
Huh 2009	8	172	18	15	0.31 [0.14, 0.52]	0.08 [0.05, 0.13]
Ittichaikulthol 2010	14	71	46	1757	0.23 [0.13, 0.36]	0.96 [0.95, 0.97]
Khan 2009a	14	64	5	297	0.74 [0.49, 0.91]	0.82 [0.78, 0.86]
Khan 2013	209	387	56	3848	0.79 [0.73, 0.84]	0.91 [0.90, 0.92]
Koh 2002	13	71	18	503	0.42 [0.25, 0.61]	0.88 [0.85, 0.90]
Konwar 2015	14	18	8	160	0.64 [0.41, 0.83]	0.90 [0.84, 0.94]
Krobbuaban 2005	36	136	33	345	0.52 [0.40, 0.64]	0.72 [0.67, 0.76]
Mehta 2014	5	11	27	433	0.16 [0.05, 0.33]	0.98 [0.96, 0.99]
Merah 2004	5	5	3	67	0.63 [0.24, 0.91]	0.93 [0.85, 0.98]
Nadal 1998	3	6	19	55	0.14 [0.03, 0.35]	0.90 [0.80, 0.96]



(*Continued . . .*)

(. . . *Continued*)

Study	TP	FP	FN	TN	Sensitivity	Specificity
Noorizad 2006	5	46	24	304	0.17 [0.06, 0.36]	0.87 [0.83, 0.90]
Qudaisat 2011	6	7	26	196	0.19 [0.07, 0.36]	0.97 [0.93, 0.99]
Salimi 2008	14	22	6	305	0.70 [0.46, 0.88]	0.93 [0.90, 0.96]
Savva 1994	11	62	6	271	0.65 [0.38, 0.86]	0.81 [0.77, 0.85]
Shah 2013	5	8	62	404	0.07 [0.02, 0.17]	0.98 [0.96, 0.99]
Singh 2009	2	6	0	292	1.00 [0.16, 1.00]	0.98 [0.96, 0.99]
Tantri 2016	5	161	23	88	0.18 [0.06, 0.37]	0.35 [0.29, 0.42]
Tse 1995	20	82	42	327	0.32 [0.21, 0.45]	0.80 [0.76, 0.84]
Vallem 2015	3	3	50	144	0.06 [0.01, 0.16]	0.98 [0.94, 1.00]
Vani 2000	2	2	6	40	0.25 [0.03, 0.65]	0.95 [0.84, 0.99]
Wajekar 2015	41	4	5	2	0.89 [0.76, 0.96]	0.33 [0.04, 0.78]
Wong 1999	7	306	0	98	1.00 [0.59, 1.00]	0.24 [0.20, 0.29]
Wong 2009	18	103	67	377	0.21 [0.13, 0.31]	0.79 [0.75, 0.82]
Yildiz 2007	23	175	57	1419	0.29 [0.19, 0.40]	0.89 [0.87, 0.91]

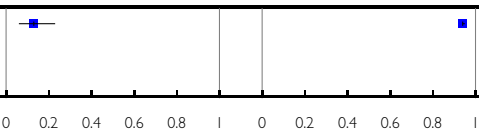


Test 10. Thyromental distance: difficult face mask ventilation.

Review: Airway physical examination tests for detection of difficult airway management in apparently normal adult patients

Test: 10 Thyromental distance: difficult face mask ventilation

Study	TP	FP	FN	TN	Sensitivity	Specificity
Kheterpal 2009	10	2955	67	50009	0.13 [0.06, 0.23]	0.94 [0.94, 0.95]

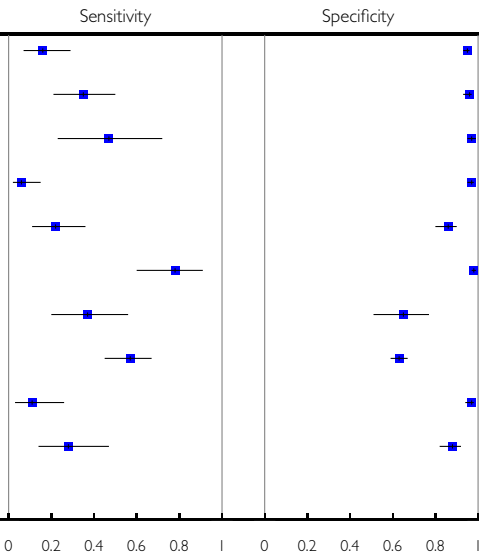


Test 11. Thyromental distance: difficult tracheal intubation.

Review: Airway physical examination tests for detection of difficult airway management in apparently normal adult patients

Test: 11 Thyromental distance: difficult tracheal intubation

Study	TP	FP	FN	TN	Sensitivity	Specificity
Arne 1998	8	60	42	1090	0.16 [0.07, 0.29]	0.95 [0.93, 0.96]
Bilgin 1998	16	20	30	434	0.35 [0.21, 0.50]	0.96 [0.93, 0.97]
Bouaggad 2004	8	8	9	295	0.47 [0.23, 0.72]	0.97 [0.95, 0.99]
Chaves 2009	4	14	61	434	0.06 [0.02, 0.15]	0.97 [0.95, 0.98]
De Jong 2015	10	34	36	202	0.22 [0.11, 0.36]	0.86 [0.80, 0.90]
K Nasa 2014	25	20	7	961	0.78 [0.60, 0.91]	0.98 [0.97, 0.99]
Knudsen 2014	11	20	19	37	0.37 [0.20, 0.56]	0.65 [0.51, 0.77]
Pottecher 1991	47	215	36	365	0.57 [0.45, 0.67]	0.63 [0.59, 0.67]
Seo 2012	4	9	32	260	0.11 [0.03, 0.26]	0.97 [0.94, 0.98]
Tuzuner-Oncul 2008	9	21	23	153	0.28 [0.14, 0.47]	0.88 [0.82, 0.92]

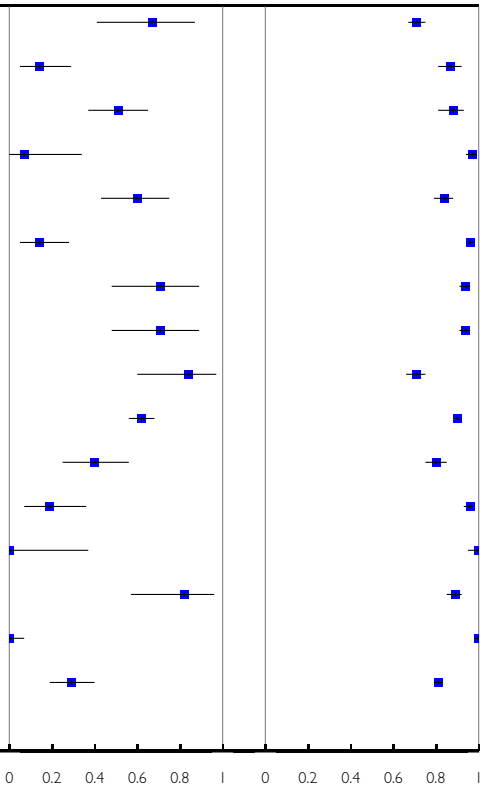


Test 12. Sternomental distance: difficult laryngoscopy.

Review: Airway physical examination tests for detection of difficult airway management in apparently normal adult patients

Test: 12 Sternomental distance: difficult laryngoscopy

Study	TP	FP	FN	TN	Sensitivity	Specificity
Al Ramadhani 1996	12	143	6	353	0.67 [0.41, 0.87]	0.71 [0.67, 0.75]
Allahyary 2008	5	22	32	144	0.14 [0.05, 0.29]	0.87 [0.81, 0.92]
Badheka 2016	26	14	25	105	0.51 [0.37, 0.65]	0.88 [0.81, 0.93]
Basaranoglu 2010	1	6	13	219	0.07 [0.00, 0.34]	0.97 [0.94, 0.99]
Basunia 2013	24	41	16	219	0.60 [0.43, 0.75]	0.84 [0.79, 0.88]
Cattano 2004	6	70	37	1843	0.14 [0.05, 0.28]	0.96 [0.95, 0.97]
Domi 2009	15	25	6	380	0.71 [0.48, 0.89]	0.94 [0.91, 0.96]
Domi 2010	15	25	6	380	0.71 [0.48, 0.89]	0.94 [0.91, 0.96]
Khan 2009a	16	106	3	255	0.84 [0.60, 0.97]	0.71 [0.66, 0.75]
Khan 2013	165	421	100	3823	0.62 [0.56, 0.68]	0.90 [0.89, 0.91]
Liaskou 2014	17	59	26	238	0.40 [0.25, 0.56]	0.80 [0.75, 0.85]
Mehta 2014	6	18	26	394	0.19 [0.07, 0.36]	0.96 [0.93, 0.97]
Merah 2004	0	0	8	72	0.0 [0.0, 0.37]	1.00 [0.95, 1.00]
Savva 1994	14	38	3	295	0.82 [0.57, 0.96]	0.89 [0.85, 0.92]
Vallem 2015	0	0	53	147	0.0 [0.0, 0.07]	1.00 [0.98, 1.00]
Yildiz 2007	23	303	57	1291	0.29 [0.19, 0.40]	0.81 [0.79, 0.83]

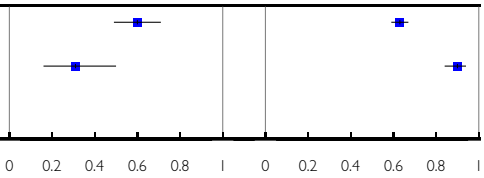


Test 13. Sternomental distance: difficult tracheal intubation.

Review: Airway physical examination tests for detection of difficult airway management in apparently normal adult patients

Test: 13 Sternomental distance: difficult tracheal intubation

Study	TP	FP	FN	TN	Sensitivity	Specificity
Pottecher 1991	50	215	33	365	0.60 [0.49, 0.71]	0.63 [0.59, 0.67]
Tuzuner-Oncul 2008	10	17	22	152	0.31 [0.16, 0.50]	0.90 [0.84, 0.94]

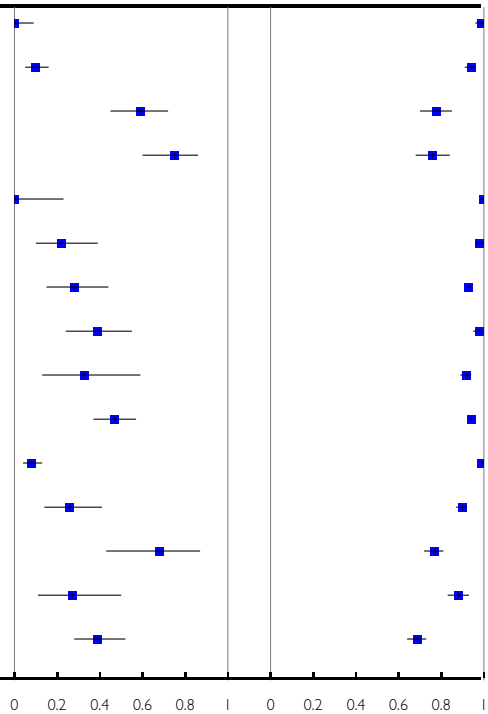


Test 14. Mouth opening: difficult laryngoscopy.

Review: Airway physical examination tests for detection of difficult airway management in apparently normal adult patients

Test: 14 Mouth opening: difficult laryngoscopy

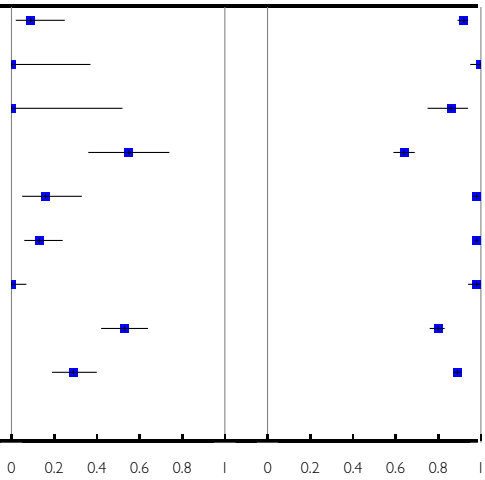
Study	TP	FP	FN	TN	Sensitivity	Specificity
Allahyary 2008	0	2	37	164	0.0 [0.0, 0.09]	0.99 [0.96, 1.00]
Ayhan 2016	12	18	111	286	0.10 [0.05, 0.16]	0.94 [0.91, 0.96]
Ayuso 2009	32	28	22	99	0.59 [0.45, 0.72]	0.78 [0.70, 0.85]
Badheka 2016	38	28	13	91	0.75 [0.60, 0.86]	0.76 [0.68, 0.84]
Basaranoglu 2010	0	0	14	225	0.0 [0.0, 0.23]	1.00 [0.98, 1.00]
Breckwoldt 2011	8	4	28	236	0.22 [0.10, 0.39]	0.98 [0.96, 1.00]
Cattano 2004	12	130	31	1783	0.28 [0.15, 0.44]	0.93 [0.92, 0.94]
Descoins 1994	16	5	25	249	0.39 [0.24, 0.55]	0.98 [0.95, 0.99]
Domi 2009	6	34	12	374	0.33 [0.13, 0.59]	0.92 [0.89, 0.94]
El-Ganzouri 1996	50	625	57	9775	0.47 [0.37, 0.57]	0.94 [0.94, 0.94]
Ezri 2003a	12	8	140	1312	0.08 [0.04, 0.13]	0.99 [0.99, 1.00]
Ezri 2003b	11	63	32	538	0.26 [0.14, 0.41]	0.90 [0.87, 0.92]
Khan 2009a	13	83	6	278	0.68 [0.43, 0.87]	0.77 [0.72, 0.81]
Konwar 2015	6	21	16	157	0.27 [0.11, 0.50]	0.88 [0.83, 0.93]
Krobbuaban 2005	27	150	42	331	0.39 [0.28, 0.52]	0.69 [0.64, 0.73]



(*Continued . . .*)

(. . . *Continued*)

Study	TP	FP	FN	TN	Sensitivity	Specificity
Mehta 2014	3	35	29	382	0.09 [0.02, 0.25]	0.92 [0.89, 0.94]
Merah 2004	0	0	8	72	0.0 [0.0, 0.37]	1.00 [0.95, 1.00]
Montemayor-Cruz 2015	0	8	5	51	0.0 [0.0, 0.52]	0.86 [0.75, 0.94]
Nasiri 2013	16	137	13	244	0.55 [0.36, 0.74]	0.64 [0.59, 0.69]
Prakash 2013	5	5	27	293	0.16 [0.05, 0.33]	0.98 [0.96, 0.99]
Shah 2013	9	7	58	407	0.13 [0.06, 0.24]	0.98 [0.97, 0.99]
Vallem 2015	0	3	53	144	0.0 [0.0, 0.07]	0.98 [0.94, 1.00]
Wong 2009	45	96	40	384	0.53 [0.42, 0.64]	0.80 [0.76, 0.83]
Yildiz 2007	23	175	57	1419	0.29 [0.19, 0.40]	0.89 [0.87, 0.91]

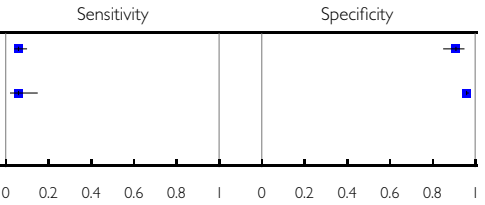


Test 15. Mouth opening: difficult face mask ventilation.

Review: Airway physical examination tests for detection of difficult airway management in apparently normal adult patients

Test: 15 Mouth opening: difficult face mask ventilation

Study	TP	FP	FN	TN	Sensitivity	Specificity
Ayhan 2016	18	12	275	122	0.06 [0.04, 0.10]	0.91 [0.85, 0.95]
Kheterpal 2009	5	1891	72	51074	0.06 [0.02, 0.15]	0.96 [0.96, 0.97]

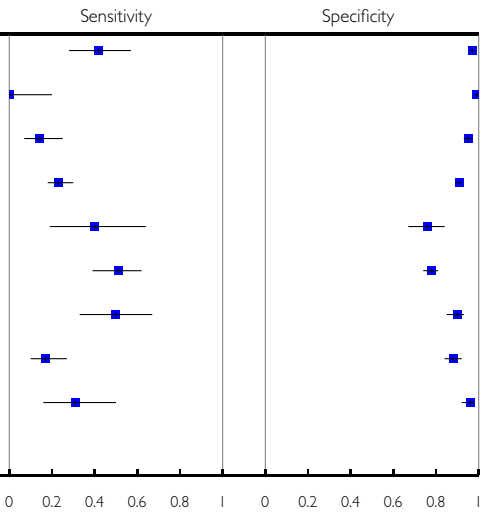


Test 16. Mouth opening: difficult tracheal intubation.

Review: Airway physical examination tests for detection of difficult airway management in apparently normal adult patients

Test: 16 Mouth opening: difficult tracheal intubation

Study	TP	FP	FN	TN	Sensitivity	Specificity
Arne 1998	21	36	29	1114	0.42 [0.28, 0.57]	0.97 [0.96, 0.98]
Bouaggad 2004	0	2	17	301	0.0 [0.0, 0.20]	0.99 [0.98, 1.00]
Chaves 2009	9	21	56	426	0.14 [0.07, 0.25]	0.95 [0.93, 0.97]
De Jong 2015	51	203	167	1964	0.23 [0.18, 0.30]	0.91 [0.89, 0.92]
Juvin 2003	8	26	12	83	0.40 [0.19, 0.64]	0.76 [0.67, 0.84]
Pottecher 1991	42	128	41	452	0.51 [0.39, 0.62]	0.78 [0.74, 0.81]
Seo 2012	18	28	18	241	0.50 [0.33, 0.67]	0.90 [0.85, 0.93]
Soyuncu 2009	15	33	71	247	0.17 [0.10, 0.27]	0.88 [0.84, 0.92]
Tuzuner-Oncul 2008	10	7	22	172	0.31 [0.16, 0.50]	0.96 [0.92, 0.98]

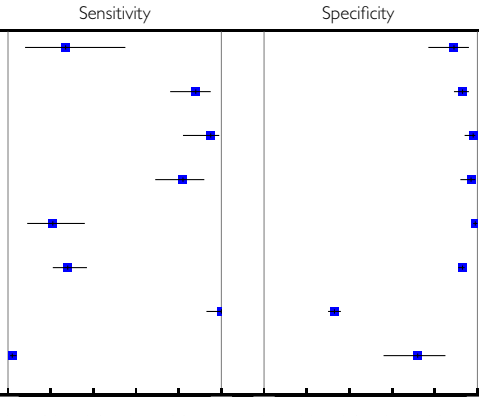


Test 17. Upper lip bite test: difficult laryngoscopy.

Review: Airway physical examination tests for detection of difficult airway management in apparently normal adult patients

Test: 17 Upper lip bite test: difficult laryngoscopy

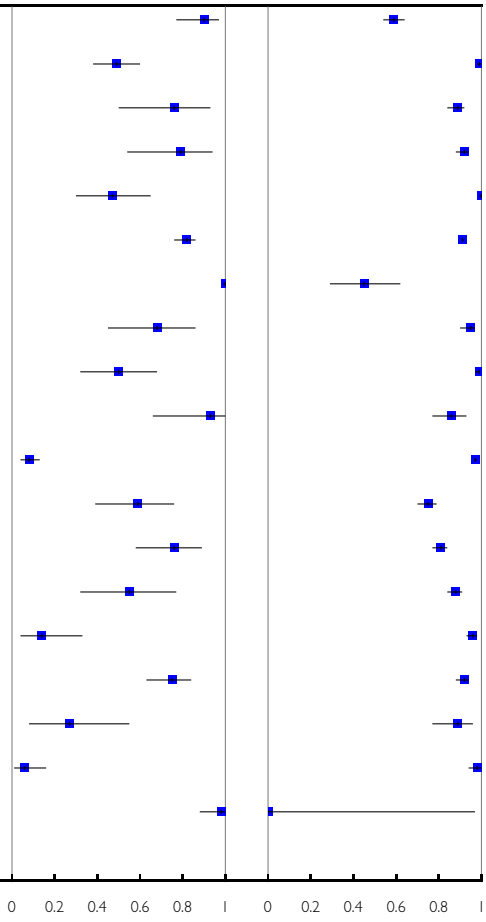
Study	TP	FP	FN	TN	Sensitivity	Specificity
Adnet 2001	4	5	11	42	0.27 [0.08, 0.55]	0.89 [0.77, 0.96]
Ali 2012	49	19	7	249	0.88 [0.76, 0.95]	0.93 [0.89, 0.96]
Allahyary 2008	35	4	2	162	0.95 [0.82, 0.99]	0.98 [0.94, 0.99]
Badheka 2016	42	4	9	115	0.82 [0.69, 0.92]	0.97 [0.92, 0.99]
Bhat 2007	8	4	31	457	0.21 [0.09, 0.36]	0.99 [0.98, 1.00]
Eberhart 2005	37	73	94	903	0.28 [0.21, 0.37]	0.93 [0.91, 0.94]
Hirmanpour 2014	53	601	0	296	1.00 [0.93, 1.00]	0.33 [0.30, 0.36]
Honarmand 2008	6	11	354	29	0.02 [0.01, 0.04]	0.73 [0.56, 0.85]



(*Continued* . . .)

(. . . *Continued*)

Study	TP	FP	FN	TN	Sensitivity	Specificity
Honarmand 2014	37	155	4	226	0.90 [0.77, 0.97]	0.59 [0.54, 0.64]
Honarmand 2015	43	3	45	504	0.49 [0.38, 0.60]	0.99 [0.98, 1.00]
Khan 2003	13	32	4	251	0.76 [0.50, 0.93]	0.89 [0.84, 0.92]
Khan 2009a	15	30	4	331	0.79 [0.54, 0.94]	0.92 [0.88, 0.94]
Khan 2011	16	0	18	266	0.47 [0.30, 0.65]	1.00 [0.99, 1.00]
Khan 2013	216	360	49	3875	0.82 [0.76, 0.86]	0.91 [0.91, 0.92]
Kolarkar 2015	260	22	0	18	1.00 [0.99, 1.00]	0.45 [0.29, 0.62]
Konwar 2015	15	10	7	178	0.68 [0.45, 0.86]	0.95 [0.90, 0.97]
Mehta 2014	16	6	16	411	0.50 [0.32, 0.68]	0.99 [0.97, 0.99]
Mishra 2009	13	12	1	74	0.93 [0.66, 1.00]	0.86 [0.77, 0.93]
Myneni 2010	14	157	159	5669	0.08 [0.04, 0.13]	0.97 [0.97, 0.98]
Nasiri 2013	17	97	12	291	0.59 [0.39, 0.76]	0.75 [0.70, 0.79]
Safavi 2014	25	85	8	359	0.76 [0.58, 0.89]	0.81 [0.77, 0.84]
Salimi 2008	11	40	9	290	0.55 [0.32, 0.77]	0.88 [0.84, 0.91]
Seo 2012	4	12	24	265	0.14 [0.04, 0.33]	0.96 [0.93, 0.98]
Shah 2013	50	35	17	378	0.75 [0.63, 0.84]	0.92 [0.88, 0.94]
Sharma 2010	4	5	11	42	0.27 [0.08, 0.55]	0.89 [0.77, 0.96]
Vallem 2015	3	3	50	144	0.06 [0.01, 0.16]	0.98 [0.94, 1.00]
Wajekar 2015	45	1	1	0	0.98 [0.88, 1.00]	0.0 [0.0, 0.97]

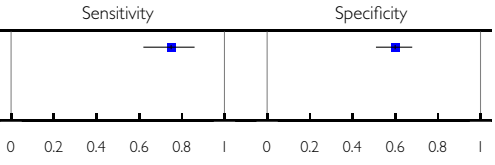


Test 18. Upper lip bite test: difficult face mask ventilation.

Review: Airway physical examination tests for detection of difficult airway management in apparently normal adult patients

Test: 18 Upper lip bite test: difficult face mask ventilation

Study	TP	FP	FN	TN	Sensitivity	Specificity
Khan 2009b	42	58	14	86	0.75 [0.62, 0.86]	0.60 [0.51, 0.68]

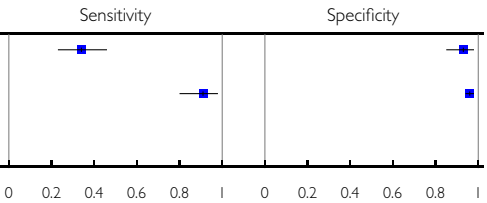


Test 19. Upper lip bite test: difficult tracheal intubation.

Review: Airway physical examination tests for detection of difficult airway management in apparently normal adult patients

Test: 19 Upper lip bite test: difficult tracheal intubation

Study	TP	FP	FN	TN	Sensitivity	Specificity
Fritscherova 2011	25	5	49	69	0.34 [0.23, 0.46]	0.93 [0.85, 0.98]
Shah 2014	43	16	4	387	0.91 [0.80, 0.98]	0.96 [0.94, 0.98]

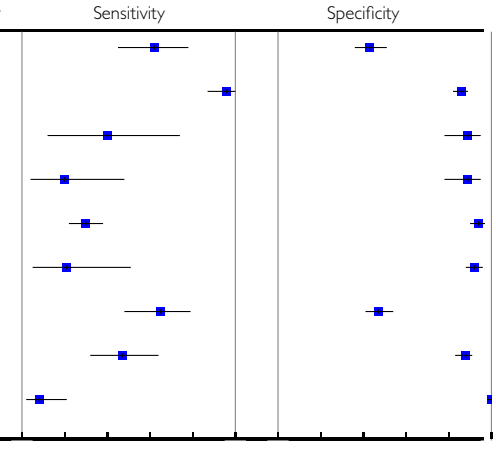


Test 20. Combination of tests: difficult laryngoscopy.

Review: Airway physical examination tests for detection of difficult airway management in apparently normal adult patients

Test: 20 Combination of tests: difficult laryngoscopy

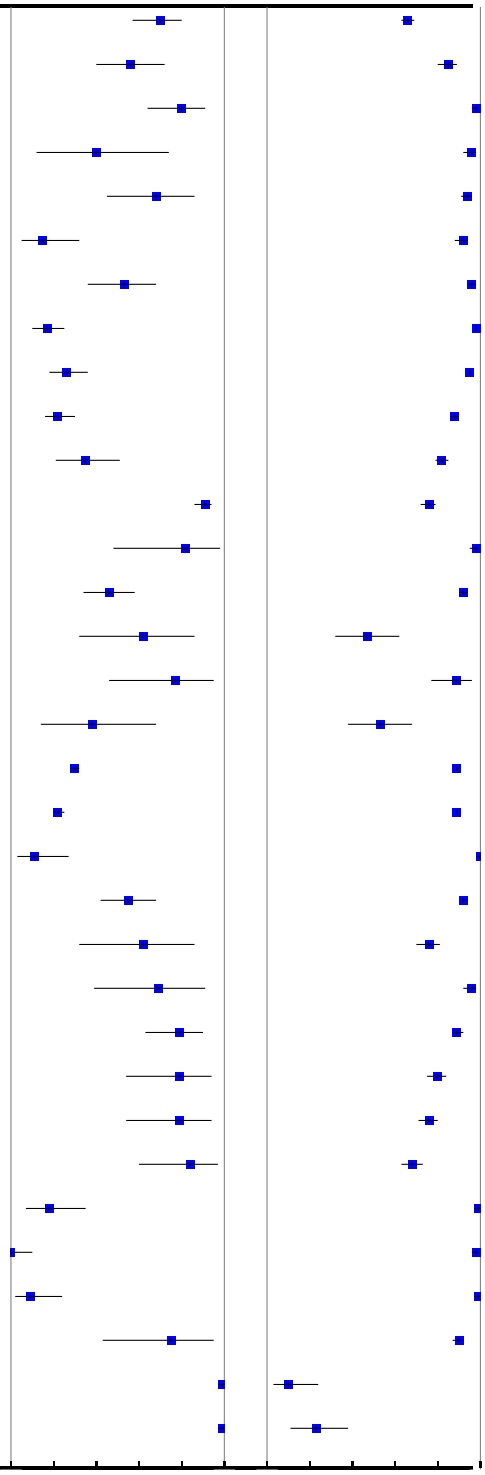
Study	TP	FP	FN	TN	Sensitivity	Specificity
Allahyary 2008	23	94	14	72	0.62 [0.45, 0.78]	0.43 [0.36, 0.51]
Ambesh 2013	51	64	2	383	0.96 [0.87, 1.00]	0.86 [0.82, 0.89]
Applegate 2013	4	7	6	56	0.40 [0.12, 0.74]	0.89 [0.78, 0.95]
Applegate 2013	3	7	12	54	0.20 [0.04, 0.48]	0.89 [0.78, 0.95]
Baig 2014	37	14	88	218	0.30 [0.22, 0.38]	0.94 [0.90, 0.97]
Basaranoglu 2010	3	17	11	208	0.21 [0.05, 0.51]	0.92 [0.88, 0.96]
Basunia 2013	26	137	14	123	0.65 [0.48, 0.79]	0.47 [0.41, 0.54]
Basunia 2013	19	32	21	228	0.48 [0.32, 0.64]	0.88 [0.83, 0.91]
Bhat 2007	3	2	36	459	0.08 [0.02, 0.21]	1.00 [0.98, 1.00]



(*Continued* . . .)

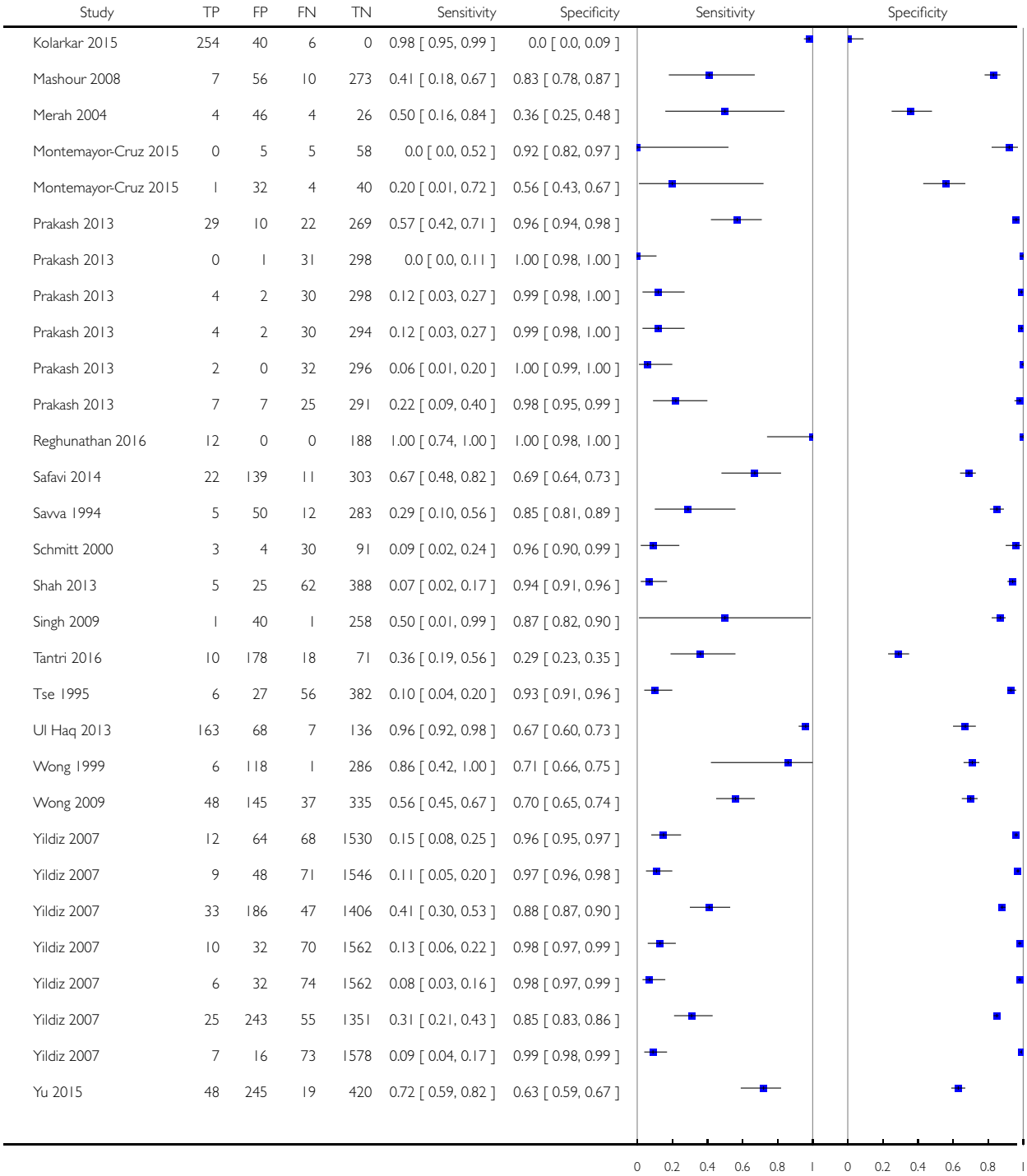
(. . . *Continued*)

Study	TP	FP	FN	TN	Sensitivity	Specificity
Cortellazzi 2007	46	313	20	615	0.70 [0.57, 0.80]	0.66 [0.63, 0.69]
Descoins 1994	23	38	18	216	0.56 [0.40, 0.72]	0.85 [0.80, 0.89]
Domi 2009	32	8	8	378	0.80 [0.64, 0.91]	0.98 [0.96, 0.99]
Domi 2009	4	10	6	220	0.40 [0.12, 0.74]	0.96 [0.92, 0.98]
Domi 2009	15	25	7	379	0.68 [0.45, 0.86]	0.94 [0.91, 0.96]
Domi 2009	5	33	28	358	0.15 [0.05, 0.32]	0.92 [0.88, 0.94]
Domi 2009	21	14	19	372	0.53 [0.36, 0.68]	0.96 [0.94, 0.98]
El-Ganzouri 1996	18	209	89	10191	0.17 [0.10, 0.25]	0.98 [0.98, 0.98]
El-Ganzouri 1996	28	487	79	9913	0.26 [0.18, 0.36]	0.95 [0.95, 0.96]
Ezri 2003a	34	156	118	1164	0.22 [0.16, 0.30]	0.88 [0.86, 0.90]
Ezri 2003b	15	107	28	494	0.35 [0.21, 0.51]	0.82 [0.79, 0.85]
Ezri 2003b	210	147	22	454	0.91 [0.86, 0.94]	0.76 [0.72, 0.79]
Frerk 1991	9	5	2	228	0.82 [0.48, 0.98]	0.98 [0.95, 0.99]
Freund 2012	32	45	38	537	0.46 [0.34, 0.58]	0.92 [0.90, 0.94]
Hashim 2014	8	25	5	22	0.62 [0.32, 0.86]	0.47 [0.32, 0.62]
Hashim 2014	10	5	3	42	0.77 [0.46, 0.95]	0.89 [0.77, 0.96]
Hashim 2014	5	22	8	25	0.38 [0.14, 0.68]	0.53 [0.38, 0.68]
Healy 2016	719	8878	1678	74877	0.30 [0.28, 0.32]	0.89 [0.89, 0.90]
Healy 2016	203	9464	707	77359	0.22 [0.20, 0.25]	0.89 [0.89, 0.89]
Honarmand 2008	4	1	31	364	0.11 [0.03, 0.27]	1.00 [0.98, 1.00]
Ittichaikulthol 2010	33	140	27	1688	0.55 [0.42, 0.68]	0.92 [0.91, 0.94]
Kalezic 2016	8	60	5	189	0.62 [0.32, 0.86]	0.76 [0.70, 0.81]
Kalezic 2016	9	11	4	238	0.69 [0.39, 0.91]	0.96 [0.92, 0.98]
Kamranmanesh 2013	30	60	8	505	0.79 [0.63, 0.90]	0.89 [0.87, 0.92]
Khan 2009a	15	73	4	288	0.79 [0.54, 0.94]	0.80 [0.75, 0.84]
Khan 2009a	15	86	4	275	0.79 [0.54, 0.94]	0.76 [0.71, 0.80]
Khan 2009a	16	116	3	245	0.84 [0.60, 0.97]	0.68 [0.63, 0.73]
Khan 2011	6	3	28	263	0.18 [0.07, 0.35]	0.99 [0.97, 1.00]
Khan 2011	0	5	34	261	0.0 [0.0, 0.10]	0.98 [0.96, 0.99]
Khan 2011	3	3	31	263	0.09 [0.02, 0.24]	0.99 [0.97, 1.00]
Khan 2014	9	60	3	516	0.75 [0.43, 0.95]	0.90 [0.87, 0.92]
Kolarkar 2015	257	36	3	4	0.99 [0.97, 1.00]	0.10 [0.03, 0.24]
Kolarkar 2015	258	31	2	9	0.99 [0.97, 1.00]	0.23 [0.11, 0.38]



(*Continued* . . .)

(. . . *Continued*)

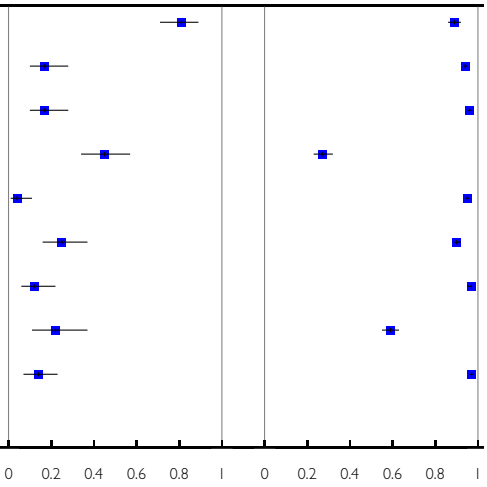


Test 21. Combination of tests: difficult face mask ventilation.

Review: Airway physical examination tests for detection of difficult airway management in apparently normal adult patients

Test: 21 Combination of tests: difficult face mask ventilation

Study	TP	FP	FN	TN	Sensitivity	Specificity
Cattano 2004	65	53	15	424	0.81 [0.71, 0.89]	0.89 [0.86, 0.92]
Langeron 2000	13	82	62	1345	0.17 [0.10, 0.28]	0.94 [0.93, 0.95]
Langeron 2000	13	56	62	1371	0.17 [0.10, 0.28]	0.96 [0.95, 0.97]
Langeron 2000	34	311	41	116	0.45 [0.34, 0.57]	0.27 [0.23, 0.32]
Langeron 2000	3	73	72	1354	0.04 [0.01, 0.11]	0.95 [0.94, 0.96]
Langeron 2000	19	137	56	1290	0.25 [0.16, 0.37]	0.90 [0.89, 0.92]
Langeron 2000	9	49	66	1378	0.12 [0.06, 0.22]	0.97 [0.95, 0.97]
Yildiz 2005	10	217	35	314	0.22 [0.11, 0.37]	0.59 [0.55, 0.63]
Yildiz 2007	11	48	69	1546	0.14 [0.07, 0.23]	0.97 [0.96, 0.98]

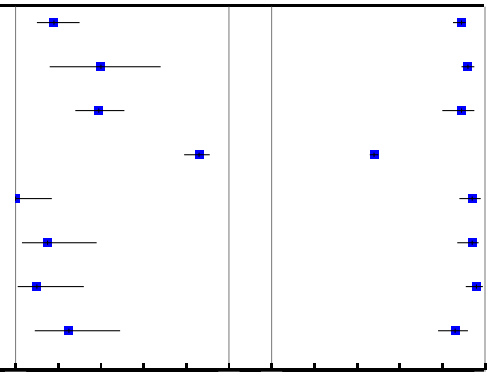


Test 22. Combination of tests: difficult tracheal intubation.

Review: Airway physical examination tests for detection of difficult airway management in apparently normal adult patients

Test: 22 Combination of tests: difficult tracheal intubation

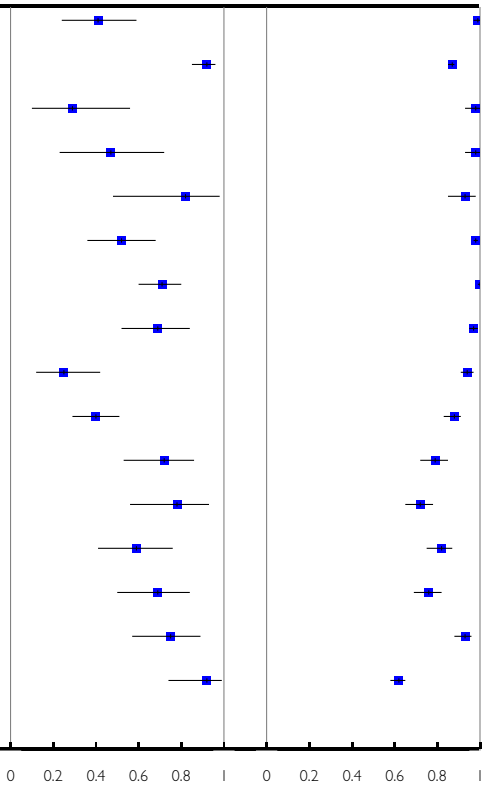
Study	TP	FP	FN	TN	Sensitivity	Specificity
Chaves 2009	12	51	53	396	0.18 [0.10, 0.30]	0.89 [0.85, 0.91]
Dohrn 2015	6	31	9	372	0.40 [0.16, 0.68]	0.92 [0.89, 0.95]
Fritscherova 2011	29	8	45	66	0.39 [0.28, 0.51]	0.89 [0.80, 0.95]
Hagiwara 2015	126	1294	21	1175	0.86 [0.79, 0.91]	0.48 [0.46, 0.50]
Juvin 2003	0	6	20	103	0.0 [0.0, 0.17]	0.94 [0.88, 0.98]
Juvin 2003	3	7	17	102	0.15 [0.03, 0.38]	0.94 [0.87, 0.97]
Juvin 2003	2	4	18	105	0.10 [0.01, 0.32]	0.96 [0.91, 0.99]
Juvin 2003	5	15	15	94	0.25 [0.09, 0.49]	0.86 [0.78, 0.92]



(*Continued . . .*)

(. . . *Continued*)

Study	TP	FP	FN	TN	Sensitivity	Specificity
K Nasa 2014	13	4	19	400	0.41 [0.24, 0.59]	0.99 [0.97, 1.00]
Kalezic 2016	101	255	9	1635	0.92 [0.85, 0.96]	0.87 [0.85, 0.88]
Kim 2011	5	2	12	104	0.29 [0.10, 0.56]	0.98 [0.93, 1.00]
Kim 2011	8	2	9	104	0.47 [0.23, 0.72]	0.98 [0.93, 1.00]
Knudsen 2014	9	5	2	70	0.82 [0.48, 0.98]	0.93 [0.85, 0.98]
Koh 2002	22	9	20	545	0.52 [0.36, 0.68]	0.98 [0.97, 0.99]
Sahin 2011	61	0	25	676	0.71 [0.60, 0.80]	1.00 [0.99, 1.00]
Seo 2012	25	7	11	262	0.69 [0.52, 0.84]	0.97 [0.95, 0.99]
Seo 2012	9	15	27	254	0.25 [0.12, 0.42]	0.94 [0.91, 0.97]
Soyuncu 2009	34	35	52	245	0.40 [0.29, 0.51]	0.88 [0.83, 0.91]
Tuzuner-Oncul 2008	23	37	9	139	0.72 [0.53, 0.86]	0.79 [0.72, 0.85]
Tuzuner-Oncul 2008	18	52	5	133	0.78 [0.56, 0.93]	0.72 [0.65, 0.78]
Tuzuner-Oncul 2008	19	32	13	144	0.59 [0.41, 0.76]	0.82 [0.75, 0.87]
Tuzuner-Oncul 2008	22	42	10	134	0.69 [0.50, 0.84]	0.76 [0.69, 0.82]
Tuzuner-Oncul 2008	24	12	8	164	0.75 [0.57, 0.89]	0.93 [0.88, 0.96]
Yu 2015	23	272	2	440	0.92 [0.74, 0.99]	0.62 [0.58, 0.65]



ADDITIONAL TABLES

Table 1. Index screening tests for the difficult airway

Test	Reference	Technique	Definition of positive response	Standard cut-off in this review
Mallampati test	Mallampati 1985	Quote: "Visibility of pharyngeal structures (faucial pillars, soft palate, and base of uvula) is noted by instructing the patient to open his/her mouth and protrude the tongue maximally while in the sitting position."	Class 1. Faucial pillars, soft palate, and uvula could be visualized Class 2. Faucial pillars and soft palate could be visualized, but uvula was masked by the base of the tongue Class 3. Only soft palate could be visualized	Class 1 and 2 versus Class 3

Table 1. Index screening tests for the difficult airway (*Continued*)

			This ordinal scale is dichotomized with assignment to Class 3 being the predictor of a DA	
Modified Mallampati test	Ezri 2001;Samsoon 1987	Quote: "All the airway assessments were done by the same anaesthesiologist, in the sitting position, with the patient's head in neutral position, mouth fully open, tongue fully extended, and without phonation. "	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 seen at all This ordinal scale is dichotomized with assignment to Class 3 and 4 being the predictor of a DA	Class 0 to 2 versus Class 3 and 4
Wilson risk score	Wilson 1988	Risk factor criteria score Weight: < 90 kg (score 0) , 90kg to 110 kg (score 1), > 110 kg (score 2) Head and neck movement: > 90 º (score 0), about 90 º (i.e. ± 10 º) (score 1), < 90 º (score 2) Jaw movement: I G ≥ 5 cm or SLux > 0 (score 0) , IG < 5 cm and SLux = 0 (score 1), I G < 5 cm and SLux < 0 (score 2) Receding mandible: normal (score 0), moderate (score 1), severe (score 2) Buck teeth: normal (score 0), moderate (score 1), severe (score 2)	The maximum possible score is 10. Higher scores are considered to be predictive of a DA. The chosen cut-off points have been > 2 or > 4	> 2
Thyromental distance	Lewis 1994	The distance between the mentum and the hyoid bone (alternatively thyroid cartilage) is measured in cm or finger widths. There is considerable variation in per-	Shorter distances are considered to be predictive of a DA. The chosen cut-off points have been < 4 cm, 6 cm, 6.5 cm, 7 cm or < 3 finger widths	6.5 cm

Table 1. Index screening tests for the difficult airway (*Continued*)

		formance of this examination. Patient position (sitting versus supine), neck position (extension versus neutral), and proximal endpoint (inside mentum versus outside mentum) are not standardized		
Sternomental distance	Ramadhani 1996	Quote: "...sternomental distance was measured as the straight distance between the upper border of the manubrium sterni and the bony point of the mentum with the head in full extension and the mouth closed. A ruler was used and the distance measured was approximated to the nearest 0.5 cm."	Shorter distances are considered to be predictive of a DA. The chosen cut-off points have been < 12.5 cm or 13.5 cm	12.5 cm
Mouth opening	Calder 2003	The interdental distance between the upper and lower incisors is measured in mm. Neck position is a factor affecting maximum mouth opening. Neck position is not standardized	Shorter distances are considered to be predictive of a DA. The chosen cut-off points have been < 3.5 cm or < 2 finger widths	3.5 cm
Upper lip bite test	Khan 2003	The patient is instructed to protrude their mandible forward and bite their upper lip	Class I. Lower incisors bite the upper lip above the vermilion border, mucosa not being visible Class II. Lower incisors bite the upper lip below the vermilion border, mucosa partially visible Class III. Lower incisors fail to bite the upper lip This ordinal scale is dichotomized with assignment to Class III being a predictor of a DA	Class I and II versus III

DA difficult airway; IG interincisor gap; SLux subluxation (maximal forward protrusion of the lower incisors beyond the upper incisors.

Table 2. Four domains for quality assessment

1	**Patient selection** **A. Risk of Bias** ***Patient sampling description*** *Signalling question 1: was a consecutive or random sample of patients enrolled?* *Signalling question 2: was a case-control design avoided?* *Signalling question 3: did the study avoid inappropriate exclusions? (Criteria met if the study did not exclude patients due to methods unusual in clinical practice, i.e. performed examination tests before study inclusion)* Signaling questions reported as yes, no, unclear ***Could the selection of patients have introduced bias?*** *Risk of bias judged as low, high, or unclear* **B. Concerns regarding applicability** Are there concerns that the included patients and setting do not match the review question? (Criteria met if the study sample did not correspond to the patient population encountered in daily clinical practice of airway management in apparently normal patients) Concerns about applicability reported as high, low, or unclear
2	**Index test** **A. Risk of bias** ***Description of index test and how it was conducted and interpreted*** *Signalling question 1: were the index test results interpreted without knowledge of the results of the reference standard? (Criteria met if index test and reference standard were conducted by different persons)* *Signalling question 2: if a threshold was used, was it prespecified?* Signalling questions reported as yes, no, unclear ***Could the conduct or interpretation of the index test have introduced bias?*** *Risk of bias judged as low, high, or unclear* **B. Concerns regarding applicability** Are there concerns that the index test, its conduct, or interpretation differed from the review question? (Applied to "non-bedside" tests, i.e. tests which require imaging techniques, etc.) Concerns about applicability reported as high, low, or unclear
3	**Reference standard** **A. Risk of bias** ***Describe condition and reference standard(s)*** *Signalling question 1: are the reference standards likely to correctly classify the target condition? (Criteria met if the study used reference standards as defined in the review)* *Signalling question 2: were the reference standards interpreted without knowledge of the results of the index test? (Criteria met if index test and reference standard were conducted by different persons)* Signalling questions reported as yes, no, unclear ***Could the reference standard, its conduct, or its interpretation have introduced bias?*** *Risk of bias judged as low, high, or unclear* **B. Concerns regarding applicability** Are there concerns that the target condition as defined by the reference standard does not match the review question? Concerns about applicability reported as high, low, or unclear
4	**Flow and timing** **A. Risk of bias** ***Describe any patients who did not receive index tests or reference standard or was excluded from 2 x 2 table*** ***Describe the interval and interventions between the index test and the reference standard*** *Signalling question 1: was there an appropriate interval between index tests and reference standard? (Usually not a problem in this*

Table 2. Four domains for quality assessment (*Continued*)

review. Considered appropriate if index tests and reference standards were conducted within a usual time-span in clinical practice, e.g. during pre-anaesthesia visit or within same hospital stay) *Signalling question 2: did all patients receive the same reference standard?* *Signalling question 3: were all patients included in the analysis?* Signalling questions reported as yes, no, unclear ***Could the patient flow have introduced bias?*** *Risk of bias judged as low, high, or unclear*

Table 3. Non-prespecified tests and combinations of screening tests for the difficult airway

Test	References	Main characteristics
Combination of ULBT and MMT	Allahyary 2008	ULBT and MMT, if any single test positive combination considered positive
Combination of MMT, TM distance, anatomical abnormalities, head movement	Ambesh 2013	MMT, TM distance, anatomical abnormalities, head movement MMT 1 to 4 points, all others 0 or 1 point > 3 points: considered positive
Telemedicine ASA checklist	Applegate 2013	ASA 11-point checklist; 2 or more points: considered positive
ASA checklist	Applegate 2013	ASA 11-point checklist; 2 or more points: considered positive
Prayer's sign	Baig 2014	Patients not able to do praying gesture considered positive
Combination of mouth opening test, TM distance, SM distance, MMT, atlanto-occipital extension	Basaranoglu 2010	Mouth opening, TM distance, SM distance, atlanto-occipital extension, MMT combination cut-off not defined
Calder test	Basunia 2013	Protrusion of lower jaw not possible: considered positive
Delilkan test	Basunia 2013	"While performing Delilkan's test the patient was asked to look straight ahead. The head was held in the neutral position. The index finger of the left hand of the observer was placed under the tip of the jaw, whereas the index finger of the right hand was placed on the patient's occipital tuberosity. The patient was now asked to look at the ceiling. If the left index finger became higher than the right, extension of the atlanto-occipital joint was considered normal."
Combination of MMT and ULBT	Bhat 2007	MMT and ULBT, if any single test positive combination considered positive
Neck mobility	Cattano 2004	Grading I to IV, III and IV: considered positive

Table 3. Non-prespecified tests and combinations of screening tests for the difficult airway (*Continued*)

Cervical mobility	Chaves 2009	< 90°: considered positive
El-Ganzouri index test	Cortellazzi 2007	Index assigning points to mouth opening, TM distance, MMT, neck movement, ability to prognath[a], body weight, history of difficult tracheal intubation > 2: considered positive
Head mobility	Descoins 1994	< 90°: considered positive
Cormack-Lehane	Dohrn 2015	III and IV: considered positive
Lower jaw protrusion	Domi 2009	Not possible: considered positive
Irregular teeth	Domi 2009	Presence of irregular teeth: considered positive
BMI	Domi 2009	> 30: considered positive
Lower jaw length	Domi 2009	< 9 cm: considered positive
Delilkan test	Domi 2009	Same definition used as Basunia 2013
Body weight	El-Ganzouri 1996	> 110 kg: considered positive
Neck movement	El-Ganzouri 1996	< 80°: considered positive
Neck movement	Ezri 2003a	< 90°: considered positive
Abnormal upper teeth	Ezri 2003b	Presence of irregular teeth: considered positive
Neck movement	Ezri 2003b	< 90°: considered positive
Combination of MMT and TM distance	Frerk 1991	MMT and TM distance, any positive considered positiveif any single test positive combination considered positive
Cormack-Lehane	Freund 2012	III and IV: considered positive
Receding mandible	Fritscherova 2011	Presence: considered positive
LEMON	Hagiwara 2015	At least one positive: considered positiveif any single item positive test considered positive
Head movement	Hashim 2014	< 35°: considered positive
Palm print sign	Hashim 2014	"Deficiency in the inter-phalangeal areas of second to fifth digit"
Prayer sign	Hashim 2014	A gap observed between the palms

Table 3. Non-prespecified tests and combinations of screening tests for the difficult airway (*Continued*)

Combination of ULBT and MMT	Healy 2016	ULBT and MMT, if any single test positive combination considered positive
Combination of MMT and TM distance	Healy 2016	MMT and TM distance, if any single test positive combination considered positive
Combination of ULBT and MMT	Honarmand 2008	ULBT and MMT, if any single test positive combination considered positive
Combination of MMT and TM distance	Ittichaikulthol 2010	MMT and TM distance, if any single test positive combination considered positive
Neck movement	Juvin 2003	< 80°: considered positive
Mandibular recession	Juvin 2003	Presence: considered positive
Abnormal teeth	Juvin 2003	Buck/missing tooth: considered positive
Hyomental distance	Kalezic 2016	< 5.3 cm: considered positive
Own score	Kalezic 2016	Including gender, age, BMI, MMT, hyomental distance
Acromioaxillosuprasternal notch index	Kamranmanesh 2013	< 0.5 considered positive
Combination of mouth opening and ULBT	Khan 2009a	Mouth opening and ULBT, if any single test positive combination considered positive
Combination of SM distance and ULBT	Khan 2009a	SM distance and ULBT, if any single test positive combination considered positive
Combination of mouth opening and SM distance	Khan 2009a	Mouth opening and SM distance, if any single test positive combination considered positive
Mandible length	Khan 2011	< 9 cm: considered positive
TM distance	Khan 2011	< 6.5 cm: considered positive
Combination of mandible length and TM distance	Khan 2011	Mandible length and TM distance, if any single test positive combination considered positive
Combination of mouth opening and ULBT	Khan 2014	Mouth opening and ULBT, if any single test positive combination considered positive
Cormack-Lehane	Kim 2011	III and IV: considered positive
Combination of Cormack-Lehane and history	Kim 2011	Cormack-Lehane and history of difficult tracheal intubation, if any single test positive combination considered positive

Table 3. Non-prespecified tests and combinations of screening tests for the difficult airway (*Continued*)

Cormack-Lehane	Knudsen 2014	III and IV: considered positive
Modified Cormack-Lehane	Koh 2002	IIb, III, IV: considered positive
Mandible length	Kolarkar 2015	< 9 cm: considered positive
Combination of mandible length and hyomental distance	Kolarkar 2015	Mandible length and hyomental distance, if any single test positive combination considered positive
Combination of mandible length and TM distance	Kolarkar 2015	Mandible length and TM distance, if any single test positive combination considered positive
Subjective anticipation	Langeron 2000	Subjective anticipation of difficult tracheal intubation by anaesthesiologist
Beard	Langeron 2000	Presence: considered positive
Lack of teeth	Langeron 2000	Lack of teeth: considered positive
Receding mandible	Langeron 2000	Presence: considered positive
Macroglossia	Langeron 2000	Presence: considered positive
Cormack-Lehane	Langeron 2000	III and IV: considered positive
Combination of ULBT and MMT	Mashour 2008	ULBT and MMT, if any single test positive combination considered positive
Mandible length	Merah 2004	< 9 cm: considered positive
Bellhouse	Montemayor-Cruz 2015	III, IV: considered positive
Patil Aldreti	Montemayor-Cruz 2015	III: considered positive
Short neck	Prakash 2013	Not defined
Mandibular protrusion	Prakash 2013	Limited protrusion: considered positive
Neck movement	Prakash 2013	< 80°: considered positive
Snoring	Prakash 2013	History of snoring: considered positive
Beard	Prakash 2013	Presence: considered positive
Receding mandible	Prakash 2013	Presence: considered positive
Own score	Reghunathan 2016	> 1.4: considered positive

Table 3. Non-prespecified tests and combinations of screening tests for the difficult airway (*Continued*)

Ratio of height to TM distance	Safavi 2014	> 29: considered positive
Jaw excursion	Sahin 2011	< 5°: considered positive
Mandibular protrusion	Savva 1994	Lack: considered positive
Neck extension	Schmitt 2000	< 80°: considered positive
Head and neck movement	Seo 2012	< 90°: considered positive
Buck teeth	Seo 2012	Presence: considered positive
Head movement	Shah 2013	< 80°: considered positive
Mandibular length	Singh 2009	< 9 cm: considered positive
Cormack-Lehane	Soyuncu 2009	III, IV: considered positive
Combination of hyomental distance and MMT	Tantri 2016	Hyomental distance and MMT, if any single test positive combination considered positive
Combination of MMT and retrognathia	Tuzuner-Oncul 2008	MMT and retrognathia, if any single test positive combination considered positive
Combination of MMT and mouth opening	Tuzuner-Oncul 2008	MMT and mouth opening, if any single test positive combination considered positive
Combination of MMT, TM distance, SM distance, and mouth opening	Tuzuner-Oncul 2008	MMT and TM distance and SM distance and mouth opening, if any single test positive combination considered positive
Combination of MMT and history	Tuzuner-Oncul 2008	MMT and history of snoring, if any single test positive combination considered positive
Cormack-Lehane	Tuzuner-Oncul 2008	III, IV: considered positive
Combination of MMT and TM distance	Tse 1995	MMT and TM distance, if any single test positive combination considered positive
Lower jaw protrusion	Ul Haq 2013	Grades A, B, C B and C: considered positive
Neck extension	K Nasa 2014	< 80°: considered positive
Combination of MMT and TM distance	Wong 1999	MMT and TM distance, if any single test positive combination considered positive

Table 3. Non-prespecified tests and combinations of screening tests for the difficult airway (*Continued*)

Mandibular luxation score	Wong 2009	Grades A, B, C B and C: considered positive
Beard	Yildiz 2005	Presence: considered positive
Mandibular protrusion	Yildiz 2007	Grades A, B, C B and C: considered positive
Combination of MMT and mandibular protrusion	Yildiz 2007	MMT and mandibular protrusion, if any single test positive combination considered positive
Combination of TM distance and mandibular protrusion	Yildiz 2007	TM distance and mandibular protrusion, if any single test positive combination considered positive
Combination of MMT and SM distance	Yildiz 2007	MMT and SM distance, if any single test positive combination considered positive
Combination of MMT and TM distance	Yildiz 2007	MMT and TM distance, if any single test positive combination considered positive
Combination of MMT and mouth opening	Yildiz 2007	MMT and mouth opening, if any single test positive combination considered positive
Combination of SM distance and mandibular protrusion	Yildiz 2007	SM distance and mandibular protrusion, if any single test positive combination considered positive
Combination of mouth opening and hyomental distance	Yildiz 2007	Mouth opening and hyomental distance, if any single test positive combination considered positive

ASA: American Society of Anesthesiologists; BMI: body mass index; MMT: modified Mallampati test; SM: sternomental; TM: thyromental; ULBT: upper lip bite test;

[a] Prognath: the ability to bring the jaw in a forward position so that the mandibular incisors are before the upper incisors.

APPENDICES

Appendix 1. Search strategy for CENTRAL, the Cochrane Library

#1 ((airway* near (test* or physical status or assess* or examinat*)) or ((distance or gap* or test* or length) near (interdental or sternomental or thyromental or interincisor* or incisor*)) or Wilson risk score or upper lip bite test or physical examin* test* or (length near upper incisor*) or (relat* and (maxillary or mandibular) and incisor*) or (visibility near uvula) or (shape near palate) or ((submandibular or mandibular) near space) or (neck near (length or thickness or diameter)) or (range and (motion or movement or flexion or extension) and (head or neck))) or mouth opening
#2 MeSH descriptor Laryngoscopy explode all trees
#3 MeSH descriptor Intubation, Intratracheal explode all trees
#4 MeSH descriptor Bronchoscopy explode all trees
#5 MeSH descriptor Laryngeal Masks explode all trees
#6 MeSH descriptor Anesthesia, this term only
#7 MeSH descriptor Laryngoscopes explode all trees
#8 (difficult* near (airway or face mask or ventilation or laryngoscopy or intubation or tracheal)) or (intubat* near (fiberoptic or stylet* or retrograde or failed)) or (laryngeal mask* or airway access):ti,ab or ((styletted or unstyletted) near tube*):ti,ab or ((laryngoscope* or Macintosh) near blade*):ti,ab or airway management:ti,ab
#9 (#2 OR #3 OR #4 OR #5 OR #6 OR #7 OR #8)
#10 (#9 AND #1)
#11 mallampati* or (difficult near intubation):ti,ab
#12 (#10 OR #11)

Appendix 2. Search strategy for MEDLINE (Ovid SP)

1. ((airway* adj3 (test* or physical status or assess* or examinat*)) or ((distance or gap* or test* or length) adj5 (interdental or sternomental or thyromental or interincisor* or incisor*)) or Wilson risk score or upper lip bite test or physical examin* test* or (length adj3 upper incisor*) or (relat* and (maxillary or mandibular) and incisor*) or (visibility adj3 uvula) or (shape adj3 palate) or ((submandibular or mandibular) adj3 space) or (neck adj3 (length or thickness or diameter)) or (range and (motion or movement or flexion or extension) and (head or neck))).mp. or mouth opening.af.
2. exp Laryngoscopy/ or exp Intubation, Intratracheal/ or exp Bronchoscopy/ or exp Laryngeal Masks/ or Anesthesia/ or Laryngoscopes/ or (difficult* adj5 (airway or face mask or ventilation or laryngoscopy or intubation or tracheal)).mp. or (intubat* adj5 (fiberoptic or stylet* or retrograde or failed)).mp. or (laryngeal mask* or airway access).ti,ab. or ((styletted or unstyletted) adj3 tube*).ti,ab. or ((laryngoscope* or Macintosh) adj3 blade*).ti,ab. or airway management.ti,ab.
3. 1 and 2
4. mallampati*.af. or (difficult adj3 intubation).ti.
5. 3 or 4

Appendix 3. Search strategy for Embase (Ovid SP)

1. ((airway* adj3 (test* or physical status or assess* or examinat*)) or ((distance or gap* or test* or length) adj5 (interdental or sternomental or thyromental or interincisor* or incisor*)) or Wilson risk score or upper lip bite test or physical examin* test* or (length adj3 upper incisor*) or (relat* and (maxillary or mandibular) and incisor*) or (visibility adj3 uvula) or (shape adj3 palate) or ((submandibular or mandibular) adj3 space) or (neck adj3 (length or thickness or diameter)) or (range and (motion or movement or flexion or extension) and (head or neck)) or mouth opening).mp.
2. exp laryngoscopy/ or exp endotracheal intubation/ or exp bronchoscopy/ or exp laryngeal mask/ or anesthesia/ or laryngoscope/ or (difficult* adj5 (airway or face mask or ventilation or laryngoscopy or intubation or tracheal)).mp. or (intubat* adj5 (fiberoptic or stylet* or retrograde or failed)).mp. or (laryngeal mask* or airway access).ti,ab. or ((styletted or unstyletted) adj3 tube*).ti,ab. or ((laryngoscope* or Macintosh) adj3 blade*).ti,ab. or airway management.ti,ab.
3. 1 and 2
4. mallampati*.af. or (difficult adj3 intubation).ti.
5. 3 or 4

Appendix 4. Search strategy for ISI Web of Science

#1 TS=(mallampati* or Wilson risk score or Upper Lip Bite test or Mouth Opening or physical examin* test*) or TS=(airway* SAME (test* or physical status or assess* or examinat*)) or TS=((distance or gap* or test* or length) SAME (sternomental or thyromental or interincisor* or incisor*)) or TS=(relat* and (maxillary or mandibular) and incisor*) or TS=(Visibility SAME uvula) or TS=(Shape SAME palate) or TS=((submandibular or mandibular) SAME space) or TS=(neck SAME (length or thickness or diameter)) or TS=(range and (motion or movement or flexion or extension) and (head or neck))
#2 TS=(endotracheal intubation or bronchoscopy or laryngeal mask) or TS=(difficult* SAME (airway or face mask or ventilation or laryngoscopy or intubation or tracheal)) or TS=(Intubat* SAME (fiberoptic or stylet* or retrograde or failed)) or TS=(laryngeal mask* or airway access) or TS=((styletted or unstyletted) SAME tube*) or TS=((laryngoscope* or Macintosh) SAME blade*) or TI=anesthesia
#3 #2 AND #1

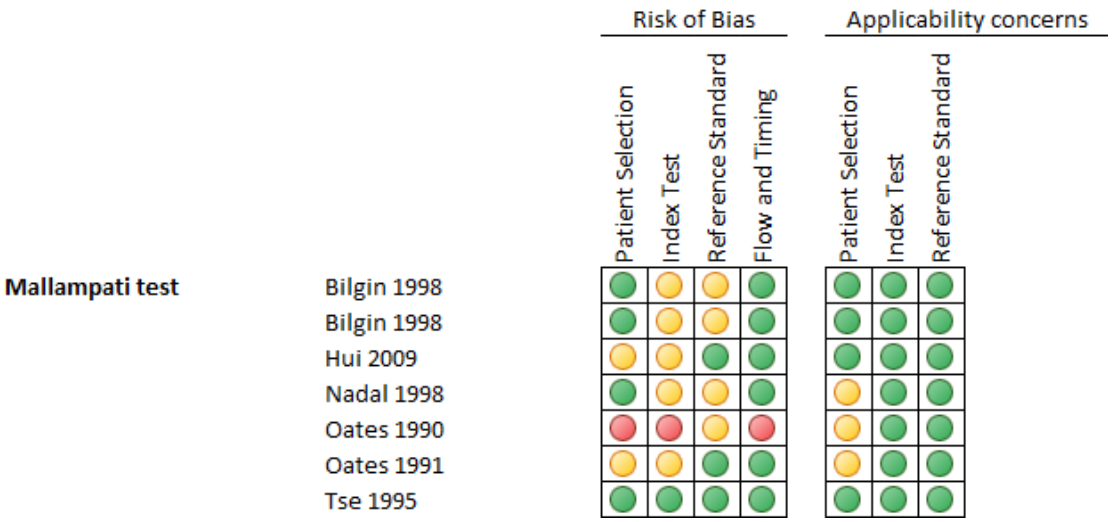
Appendix 5. Search strategy for CINAHL (EBSCO host)

S1 (MM "Physical Examination")
S2 TX mallampati* or TX Wilson risk score or TX Upper Lip Bite test or TX Mouth Opening or TX physical examin* test*
S3 airway* N3 (test* or physical status or assess* or examinat*)
S4 ((distance or gap* or test* or length) N4 (sternomental or thyromental or interincisor* or incisor*))
S5 Length N3 upper incisor*
S6 relat* and (maxillary or mandibular) and incisor*
S7 Visibility N3 uvula
S8 Shape N3 palate
S9 ((submandibular or mandibular) N3 space)
S10 (neck N3 (length or thickness or diameter))
S11 (range and (motion or movement or flexion or extension) and (head or neck))
S12 S1 or S2 or S3 or S4 or S5 or S6 or S7 or S8 or S9 or S10 or S11
S13 (MM "Laryngoscopy") OR (MM "Intubation, Intratracheal") OR (MM "Bronchoscopy") OR (MM "Laryngeal Masks") OR (MH "Anesthesia+")
S14 difficult* N4 (airway or face mask or ventilation or laryngoscopy or intubation or tracheal)
S15 Intubat* N4 (fiberoptic or stylet* or retrograde or failed)
S16 AB laryngeal mask* or AB airway access
S17 ((styletted or unstyletted) N3 tube*)
S18 ((laryngoscope* or Macintosh) N3 blade*)
S19 S13 or S14 or S15 or S16 or S17 or S18
S20 S19 and S12

Appendix 6. Risk of bias and applicability concerns summary figures

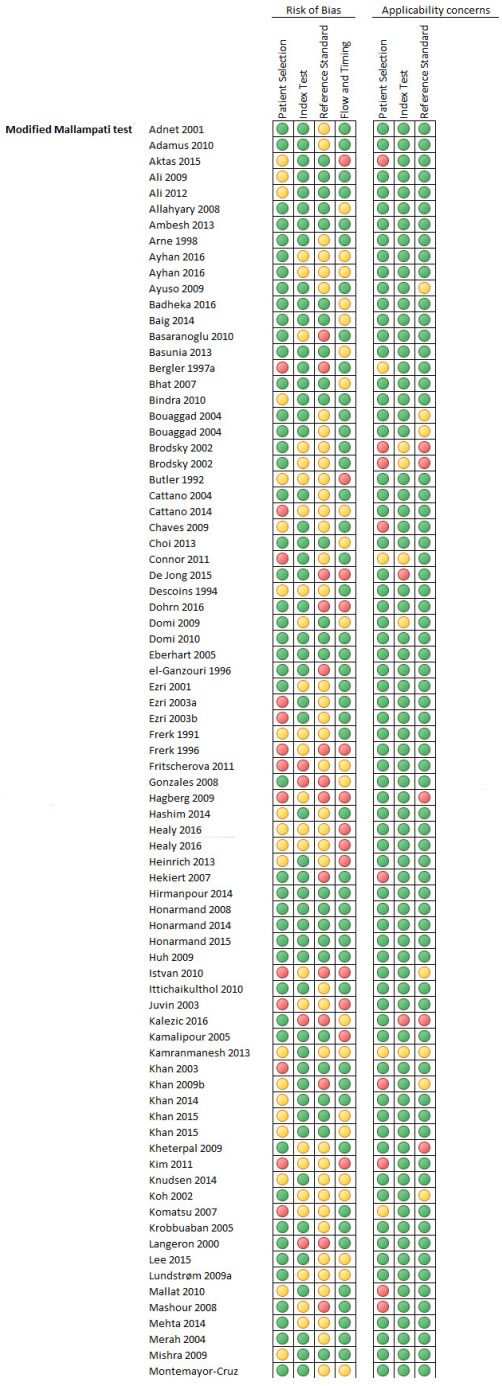
Mallampati test Figure 16

Figure 16. Risk of bias and applicability concerns summary for Mallampati test: review authors' judgements about each domain for each included study.



Modified Mallampati test (part 1) Figure 17

Figure 17. Risk of bias and applicability concerns summary for modified Mallampati test (part 1): review authors' judgements about each domain for each included study.

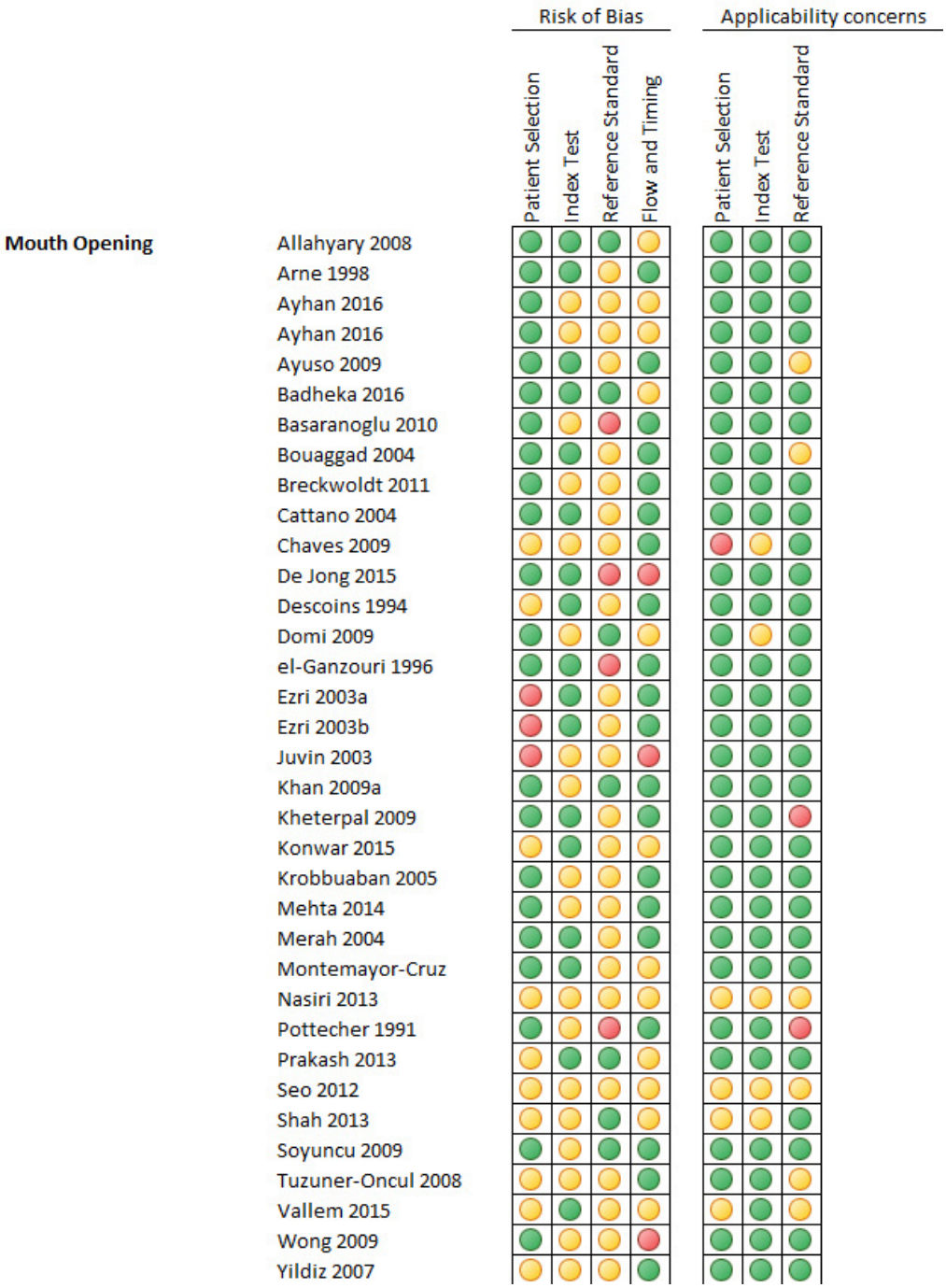


Modified Mallampati test (part 2) Figure 18

Figure 18. Risk of bias and applicability concerns summary for modified Mallampati test (part 2): review authors' judgements about each domain for each included study.

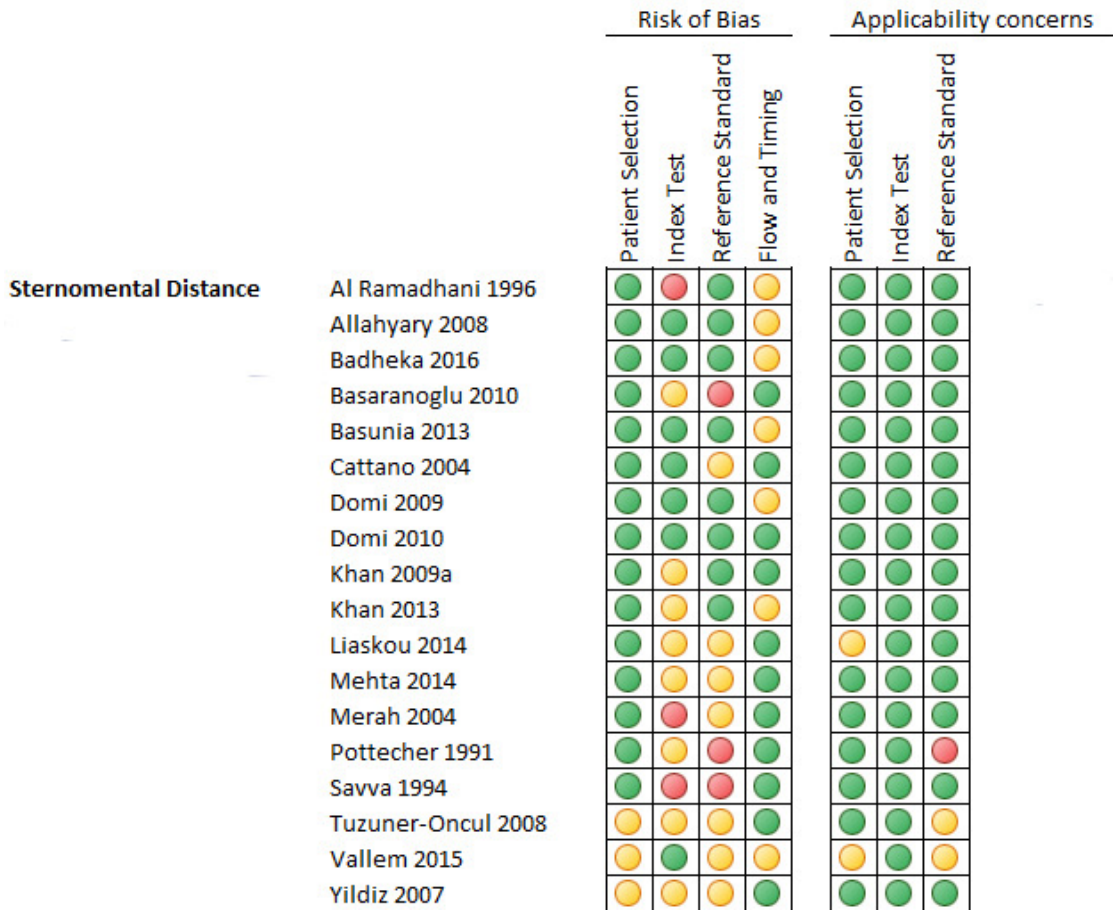
Mouth opening test Figure 19

Figure 19. Risk of bias and applicability concerns summary for mouth opening: review authors' judgements about each domain for each included study.



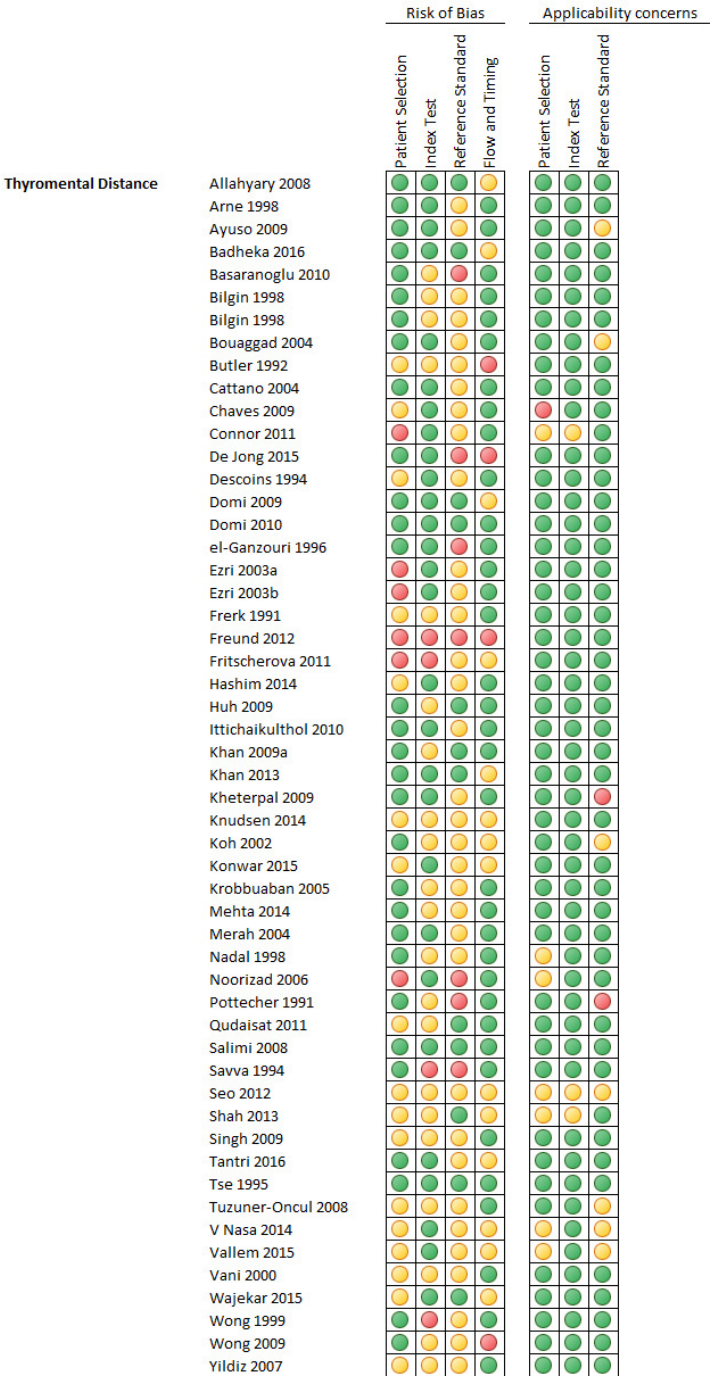
Sternomental distance Figure 20

Figure 20. Risk of bias and applicability concerns summary for sternomental distance: review authors' judgements about each domain for each included study.



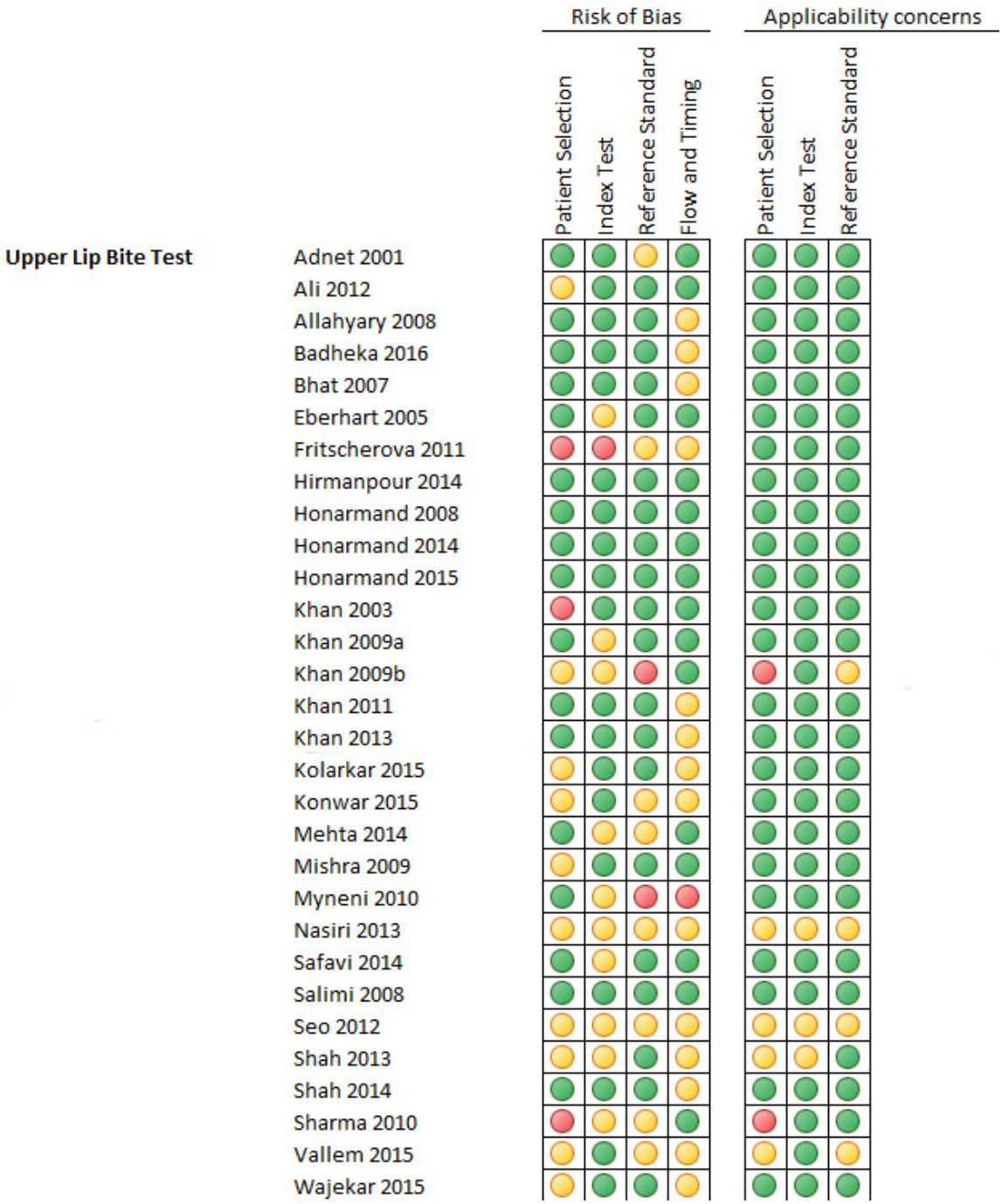
Thyromental distance Figure 21

Figure 21. Risk of bias and applicability concerns summary for thyromental distance: review authors' judgements about each domain for each included study.



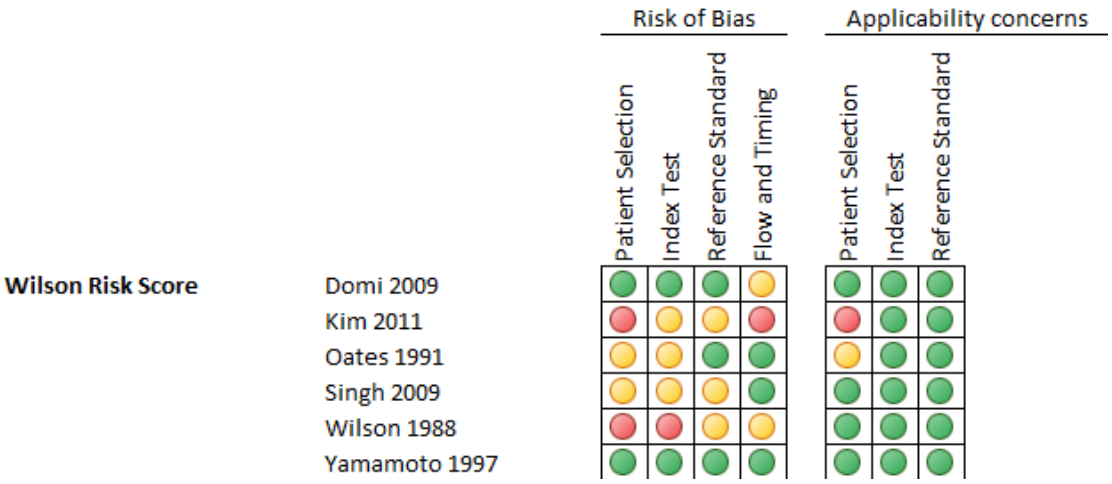
Upper lip bite test Figure 22

Figure 22. Risk of bias and applicability concerns summary for upper lip bite test: review authors' judgements about each domain for each included study.



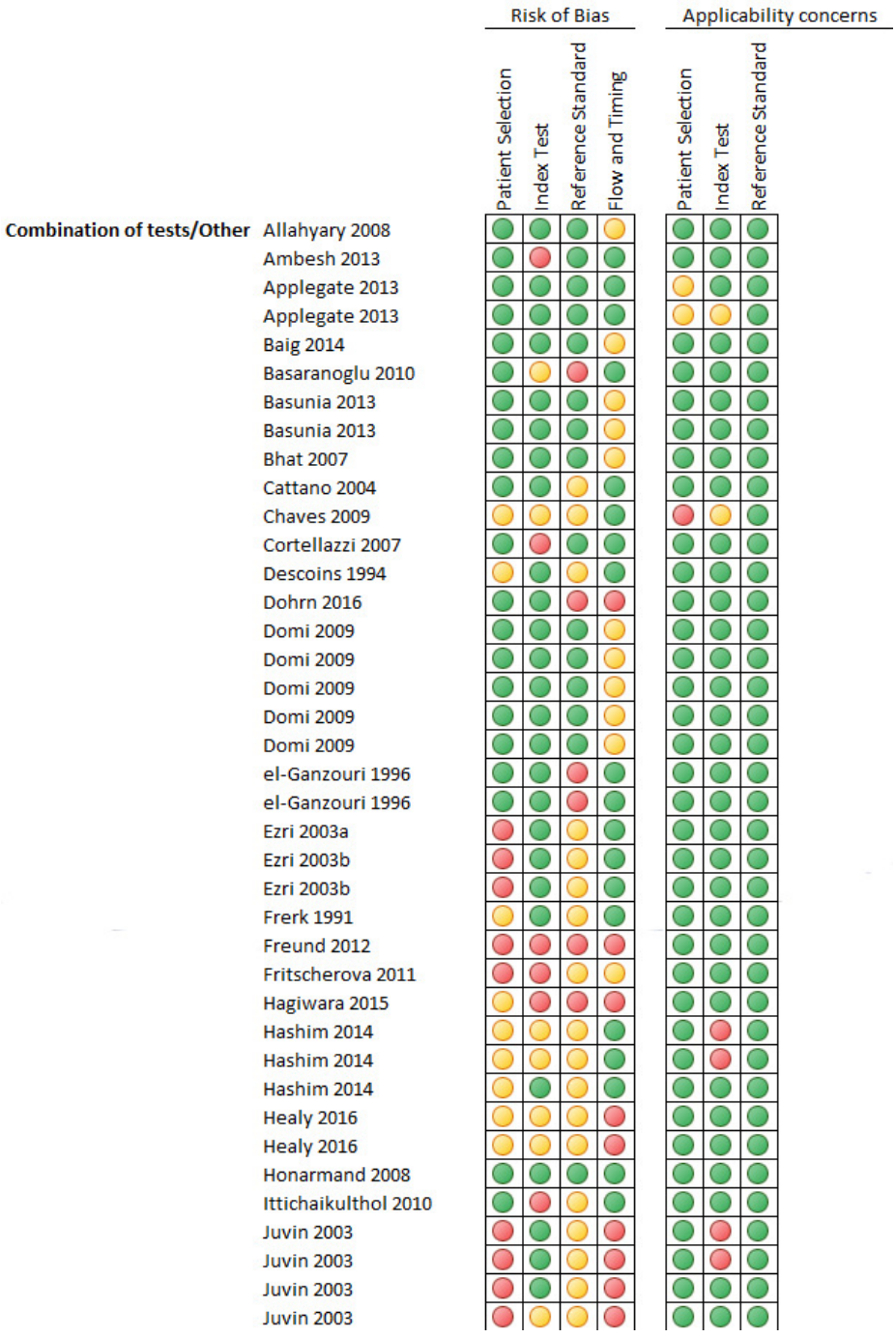
Wilson risk score Figure 23

Figure 23. Risk of bias and applicability concerns summary for Wilson risk score: review authors' judgements about each domain for each included study.



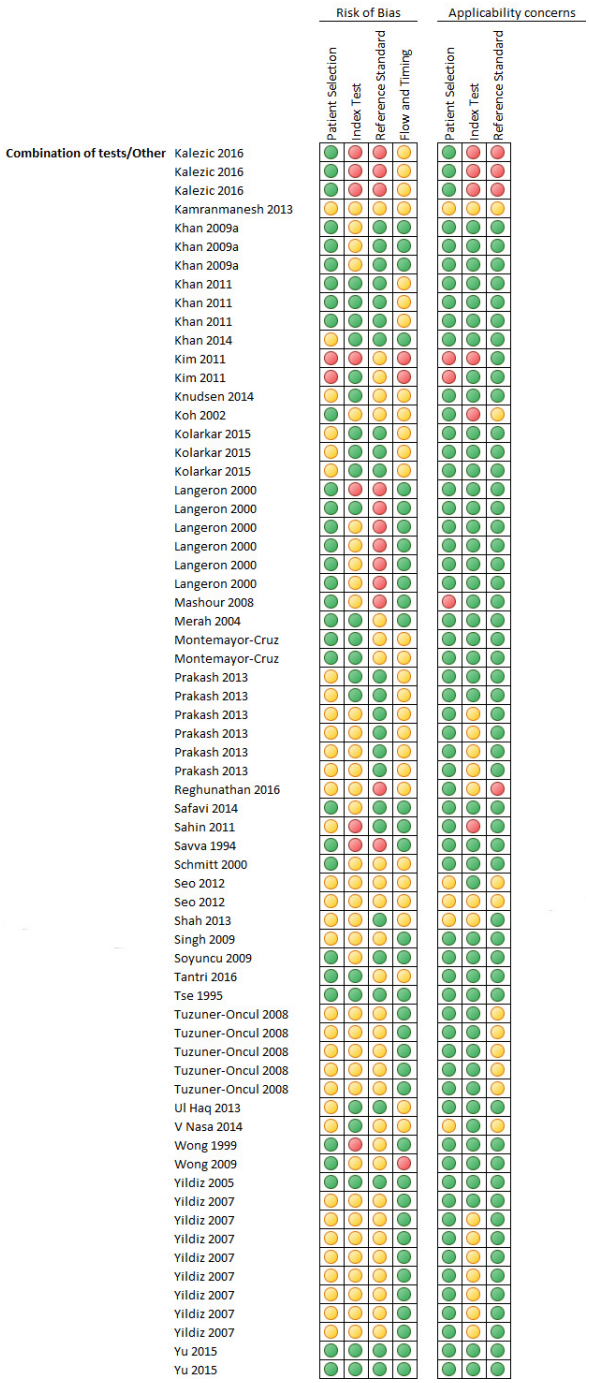
Combinations of tests (part 1) Figure 24

Figure 24. Risk of bias and applicability concerns summary for combination of tests (part 1): review authors' judgements about each domain for each included study.



Combinations of tests (part 2) Figure 25

Figure 25. Risk of bias and applicability concerns summary for combination of tests (part 2): review authors' judgements about each domain for each included study.



CONTRIBUTIONS OF AUTHORS

Conceiving the review: Nathan L Pace (NLP)

Designing the review: Anna Lee (AL), NLP, Harald Herkner (HH), Karen Hovhannisyan (KH)

Co-ordinating the review: HH, NLP

Undertaking manual searches: KH, HH, Dominik Roth (DR)

Screening search results: DR, NLP, HH, AL

Organizing retrieval of papers: AL, KH, DR, HH

Screening retrieved papers against inclusion criteria: NLP, HH, DR, AL

Appraising quality of papers: NLP, HH, Alexandra Warenits (AW), Jasmin Arrich (JA), DR, AL

Abstracting data from papers: NLP, HH, AW, JA, DR, AL

Writing to authors of papers for additional information: DR, KH, JA

Providing additional data about papers: AL, KH, JA

Obtaining and screening data on unpublished studies: DR, KH

Data management for the review: NLP, HH, DR

Entering data into Review Manager (Review Manager 2014): NLP, HH, DR

Checking data entry in Review Manager (Review Manager 2014): AL, DR

Review Manager statistical data (Review Manager 2014): AL, NLP, HH, DR

Other statistical analysis, not using Review Manager (Review Manager 2014): NLP

Interpretation of data: AL, NLP, HH, DR

Statistical inferences: AL, NLP, HH, DR

Writing the review: AL, NLP, HH , KH, DR, AW, JA

Providing guidance on the review: AL, NLP, HH, KH, DR

Securing funding for the review: not applicable

Performing previous work that was the foundation of the present study: AL

Guarantor for the review (one author): HH

Person responsible for reading and checking review before submission: AL, NLP, HH, KH, AW, JA, DR

DECLARATIONS OF INTEREST

Dominik Roth: none known

Nathan L Pace: none known

Anna Lee: is the first author of a previously published diagnostic test accuracy review of the Mallampati score (Lee 2006).

Karen Hovhannisyan: none known

Alexandra-Maria Warenits: none known

Jasmin Arrich: none known

Harald Herkner: none known

This review was selected for the third Cochrane Review Support Programme.

SOURCES OF SUPPORT

Internal sources

- The Chinese University of Hong Kong, Shatin, NT, Hong Kong.
- Medical University of Vienna, Vienna, Austria.
- The Cochrane Anaesthesia Review Group, Rigshospitalet, Copenhagen, Denmark.
- University of Utah, Salt Lake City, UT, USA.
- Third Cochrane Review Support Programme, Other.

External sources

- No sources of support supplied

DIFFERENCES BETWEEN PROTOCOL AND REVIEW

We removed the secondary objective of this review: to determine which test or combination of tests has the highest accuracy in studies with direct comparisons for assessing the physical status of the airway in patients with no apparent anatomical airway abnormalities.