BJA

Predictive value of the El-Ganzouri multivariate risk index for difficult tracheal intubation: a comparison of Glidescope® videolaryngoscopy and conventional Macintosh laryngoscopy

P. Cortellazzi[1]*, L. Minati[2], C. Falcone[3], M. Lamperti[1] and D. Caldiroli[1]

[1]*Neuroanaesthesia Department,* [2]*Scientific Direction Unit and* [3]*Neuroepidemiology Department, Fondazione Istituto Neurologico Carlo Besta IRCCS, Via Celoria, 11, I-20133 Milano, MI, Italy*

*Corresponding author. E-mail: pcortellazzi@istituto-besta.it; lminati@istituto-besta.it

Background. The predictive value of the El-Ganzouri risk index (EGRI) for difficult intubation has been evaluated using Macintosh laryngoscopy as reference standard. The Glidescope® videolaryngoscope provides improved visualization of the glottis. We studied the predictive value of the EGRI using videolaryngoscopy as reference standard.

Methods. Data from two subsequent groups of patients, intubated with Macintosh laryngoscopy (ML, $n=994$) and videolaryngoscopy (VL, $n=843$), were retrospectively analysed. The EGRI was taken as index test. The two types of laryngoscopy were adopted as reference for the presence of Cormack and Lehane grading III–IV. For both groups, sensitivity, specificity, and positive and negative post-test probabilities (PTP) were calculated for thresholds on the EGRI scale. Receiver operating characteristic curves and corresponding areas (AUC) were obtained.

Results. Sensitivity and specificity were 69.7% and 66.3% at the cut-off value of 2 in the ML group, and 93.3% and 76.6% at the cut-off value of 3 in the VL group. Corresponding positive and negative PTP were 12.81% and 3.15% in the ML group, and 6.73% and 0.16% in the VL group. At the threshold of 4, positive and negative PTP were 31.34% and 4.85% in the ML group. At the threshold of 7, positive and negative PTP were 85.71% and 1.08% in the VL group. The AUC was 0.74 in the ML group and 0.91 in the VL group.

Conclusions. The predictive value of the EGRI may have been underestimated due to limited accuracy of Macintosh laryngoscopy. Using videolaryngoscopy, the EGRI might be reconsidered as a decisional tool.

Br J Anaesth 2007; **99**: 906–11

Keywords: complications, intubation tracheal; equipment, laryngoscopes; larynx, laryngoscopy; risk

Accepted for publication: September 5, 2007

Predictive indexes for difficult intubation aim to warn of difficult laryngeal exposure and tracheal intubation when risk factors are not evident, enabling safer strategies to be adopted.[1–4] The multivariate risk index developed by El-Ganzouri and colleagues combines and stratifies seven variables derived from parameters and observations individually associated with difficult intubation.[3] Evaluation of the predictive value of the El-Ganzouri risk index (EGRI) has been performed using conventional Macintosh laryngoscopy as reference standard.[3]

In recent years, several alternatives to Macintosh laryngoscopy have been proposed, aiming at obtaining better visualization of the laryngeal structures. As depicted in Figure 1A, the Glidescope® videolaryngoscope, which has been available since 2003, consists of a handle, similar to that of a standard laryngoscope, and a non-detachable blade, which has a maximum width of 18 mm and a curvature of 60° in the midline.[5–8] A digital camera and two light-emitting diodes are embedded at the tip of the blade. The wide-angle lens, central insertion of the blade, and the fact that the camera always remains remote from the laryngeal structures result in a rather wide visual field, as exemplified in Figure 1B.[5–8]
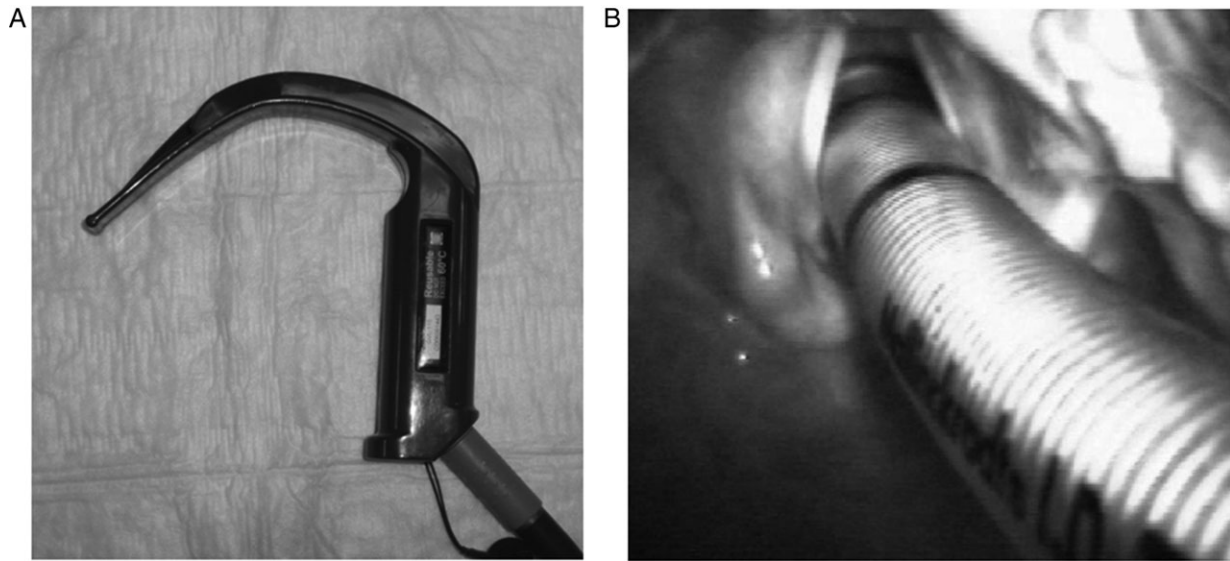
Fig 1 Lateral profile of the Glidescope® blade (A) and view of the laryngeal structures after introduction of the tube into the trachea (B).

Early studies with the Glidescope® videolaryngoscope reported that it may make intubation easier when difficulty is caused by insufficient exposure of the larynx.[5–8] The purpose of this study was to determine the predictive value of the El-Ganzouri index using videolaryngoscopy in place of conventional laryngoscopy as reference standard.

Methods

This study is based on retrospective analysis of anaesthesiological records of patients undergoing brain or spine surgery from 2004 to 2006. It was conducted in compliance with institutional guidelines for research and data protection, and in accordance with the principles of the Helsinki declaration.

In 2004, in our institution, patients were routinely intubated by means of Macintosh laryngoscopy. In 2005, Macintosh laryngoscopy was replaced by Glidescope® videolaryngoscopy. During the transitional period, each member of the team of 12 anaesthesiologists (unchanged since 2002) performed at least 50 intubations under supervision of an anaesthesiologist already familiar with the technique. This ensured that after the training period, intubation difficulties could not be caused by incomplete acquisition of a level of skill corresponding to the indications of Cuchillo and Rodriguez[9] and Kramer and Osborn.[10] In 2006, patients were routinely intubated by means of Glidescope® videolaryngoscopy.

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. The group of patients intubated with Macintosh laryngoscopy (ML group) included 994 entries, and those intubated with Glidescope® videolaryngoscopy (VL group) included 843 entries. The characteristics of

the patients in the two groups, namely age, gender, and BMI, are reported in Table 1.

Patients younger than 16 yr, those with known airway pathology or cervical spine injury, those who required rapid sequence induction, those for whom the El-Ganzouri index was not available (31 patients overall), and those for whom it had been decided to perform awake intubation with flexible fibreoptic bronchoscopy (one in the ML group and one in the VL group) were excluded from the study.

The El-Ganzouri index, ranging from 0 to 12, defined in Table 2, has been in routine use in our department since

Table 1 Characteristics, El-Ganzouri scores, outcomes, and number of failed intubations for patients intubated with Macintosh largyngoscopy (ML group) and patients intubated with Glidescope® videolaryngoscopy (VL group). Age and BMI are given as mean (range) mean (SD), respectively. *Statistically significant differences

	ML group (n = 994)	VL group (n = 843)
Age (yr)	50.1 (16–83)	51.5 (16–84)
Sex		
Male	517 (52.0%)	435 (51.6%)
Female	477 (48.0%)	408 (48.4%)
BMI (kg m^{-2})	25.3 (5.0)	26.0 (4.7)
El-Ganzouri index*		
0	339 (34.1%)	240 (28.5%)
1	296 (29.8%)	239 (28.4%)
2	187 (18.8%)	156 (18.4%)
3	105 (10.6%)	105 (12.5%)
4	38 (3.8%)	59 (7.0%)
5	22 (2.2%)	22 (2.6%)
6	5 (0.5%)	15 (1.8%)
7	1 (0.1%)	1 (0.1%)
8	0 (0.0%)	2 (0.2%)
9	0 (0.0%)	3 (0.4%)
10	1 (0.1%)	1 (0.1%)
Outcome*		
Favourable (F)	928 (93.4%)	828 (98.2%)
Unfavourable (U)	66 (6.6%)	15 (1.8%)
Failed intubations	3 (0.3%)	0 (0%)

Table 2 Composition of the El-Ganzouri and colleagues[3] multivariate risk index

Variable	Score	Variable	Score
Mouth opening		Ability to prognath	
≥4 cm	+0	Yes	+0
<4 cm	+1	No	+1
Tyromental distance		Body weight	
>6.5 cm	+0	<90 kg	+0
6.0–6.5 cm	+1	90–110 kg	+1
<6.0 cm	+2	>110 kg	+2
Modified Mallampati class		History of difficult intubation	
I (soft palate, fauces, uvula, and pillars seen)	+0	None	+0
II (soft palate, fauces, and uvula seen)	+1	Questionable	+1
III (soft palate, base of uvula seen)	+2	Definite	+2
IV (soft palate not visible)	+2		
Neck movement			
>90°	+0		
80–90°	+1		
<80°	+2		

2002 and is recorded 1 day before surgery. The modified Mallampati test is performed, giving a score between 1 and 4 as defined in Table 2.[11] This test is performed in the sitting position without phonation. In general, the anaesthesiologist measuring the index was different from the one who performed laryngeal exposure and tracheal intubation.

One hour before induction, all patients received i.m. atropine 8 μg kg^{-1}. All patients were connected to standard monitoring devices and received i.v. sodium thiopental 4–6 mg kg^{-1} or propofol 1–2 mg kg^{-1}, and remifentanil 0.05–0.15 μg kg^{-1} min^{-1}. Neuromuscular block was achieved using vecuronium 0.1 mg kg^{-1}. In both groups, patients were placed in the 'sniffing' position with their head on a pillow.

The degree of laryngeal exposure was scored, with direct Macintosh laryngoscopy (ML group) or with Glidescope® videolaryngoscopy (VL group), according to the Cormack and Lehane (C&L) grading system. A graphical representation, similar to that introduced by Samsoon and Young,[11] was printed on the anaesthesiological record for guidance.

After determining the C&L grade, the 'BURP' manoeuvre (i.e. displacement of the larynx by backward, upward, and rightward pressure on the thyroid cartilage) was performed in both groups if laryngeal exposure was judged insufficient.[12] In the ML group, Frova's tracheal introducer was used if intubation was difficult.[13] In the VL group, the tube was routinely preformed by means of a 'hockey stick'-shaped flexible stylet.[10] The former is meant for use with traditional laryngoscopy only, whereas the latter facilitates introduction of the tube with the blade design of the Glidescope® videolaryngoscope.[10 13]

We adopted a strict rule not to perform more than three intubation attempts; upon failure of the third attempt, infusion of remifentanil was suspended, prostigmine 30 μg kg^{-1} and atropine 20 μg kg^{-1} were administered i.v., and the patient was ventilated by means of bag-mask or laryngeal mask until spontaneous breathing resumed.

For the purpose of this study, C&L scores I and II were considered as a favourable outcome (F), whereas C&L scores III and IV were considered as an unfavourable outcome (U).

The ML and VL groups were compared using unpaired *t*-tests for age and BMI and χ^2 tests for gender, outcome (F/U), distribution of El-Ganzouri scores, and proportion of cases in which the same anaesthesiologist measured the index and performed laryngeal exposure. The period prevalence of unfavourable outcome was computed for both groups.[14] A *P*-value of <0.05 was considered as statistically significant.

For both groups and for each score on the El-Ganzouri scale, the number of true positives, false positives, true negatives, and false negatives with respect to outcome type (F or U) was determined, and values of sensitivity and specificity were computed. For each threshold between 1 and 9 on the El-Ganzouri scale, positive and negative likelihood ratios (LR+ and LR−) and corresponding 95% confidence intervals, and positive and negative post-test probabilities (PTP) [P(D+|T+) and P(D+|T−)] were calculated. The positive PTP corresponds to the probability of unfavourable outcome with a score equal to or greater than the one under consideration. The negative PTP corresponds to the probability of unfavourable outcome with a score less than the one under consideration. Unlike sensitivity and specificity and likelihood ratios, positive and negative PTP also depend on period prevalence.[15] For both groups, the cut-off value was determined as the score on the El-Ganzouri scale which had the effect of maximizing the sum of sensitivity and specificity.

Receiver operating characteristic (ROC) curves were also obtained for both groups. Here, the proportion of true positives (i.e. sensitivity) was plotted against the proportion of false positives (i.e. 1-specificity), for each cut-off value. The area under curve (AUC) was then calculated, and the presence of significant differences with respect to AUC=0.5 (corresponding to an uninformative test) was tested for by means of a *z*-test computed with $(AUC-0.5)/\sqrt{(\sigma^2_{AUC})}$.

Results

There were no significant differences in age, BMI, and gender between the ML and VL groups. The distribution of scores on the El-Ganzouri scale is given in Table 1. The El-Ganzouri score is higher in the VL group (*P*=0.02). The maximum value observed was 10. The anaesthesiologist measuring the index and performing laryngeal exposure was the same in 80 cases (8%) in the ML group and in 84 cases (10%) in the VL group; the proportion was not significantly different between the two groups. The period prevalence of unfavourable outcome (C&L grade III–IV) in the two groups is reported in

Table 3 Number of true positives (TP), true negatives (TN), false positives (FP) and false negatives (FN), and sensitivity (SE), specificity (SP), positive and negative likelihood ratios (LR+ and LR−, with 95% confidence intervals), and PTP [P(D+|T+) and P(D+|T−)] for thresholds on the EGRI, for the two groups. *The cut-off values, maximizing the sum of sensitivity and specificity

EGRI	TP	TN	FP	FN	SE (%)	SP (%)	LR+	LR−	P(D+|T+) (%)	P(D+|T−) (%)
MacIntosh laryngoscopy (ML group)										
0	66	0	928	0	100	0				
1	59	332	596	7	89.4	35.8	1.39 (1.26–1.53)	0.30 (0.15–0.60)	9.01	2.06
2*	46	615	313	20	69.7	66.3	2.07 (1.72–2.48)	0.46 (0.32–0.66)	12.81	3.15
3	31	787	141	35	47.0	84.8	3.09 (2.29–4.16)	0.63 (0.50–0.79)	18.02	4.26
4	21	882	46	45	31.8	95.0	6.42 (4.09–10.08)	0.72 (0.61–0.85)	31.34	4.85
5	12	911	17	54	18.2	98.2	9.93 (4.95–19.90)	0.83 (0.74–0.93)	41.38	5.60
6	4	925	3	62	6.1	99.7	18.75 (4.28–82.03)	0.94 (0.89–1.00)	57.14	6.28
7	1	927	1	65	1.5	99.9	14.06 (0.89–222.29)	0.99 (0.96–1.02)	50.00	6.55
8	1	927	1	65	1.5	99.9	14.06 (0.89–222.29)	0.99 (0.96–1.02)	50.00	6.55
9	1	927	1	65	1.5	99.9	14.06 (0.89–222.29)	0.99 (0.96–1.02)	50.00	6.55
10	0	927	1	66	0	99.9				
Glidescope® videolaryngoscopy (VL group)										
0	15	0	828	0	100	0				
1	15	240	588	0	100	29	1.36 (1.24–1.51)	0.11 (0.01–1.65)	2.41	0.19
2	15	479	349	0	100	57.9	2.30 (2.04–2.59)	0.05 (0.01–0.83)	4.00	0.10
3*	14	634	194	1	93.3	76.6	3.98 (3.32–4.78)	0.09 (0.01–0.58)	6.73	0.16
4	9	734	94	6	60	88.6	5.29 (3.35–8.33)	0.45 (0.24–0.84)	8.74	0.81
5	7	791	37	8	46.7	95.5	10.44 (5.58–19.53)	0.56 (0.35–0.90)	15.91	1.00
6	6	812	16	9	40	98.1	20.70 (9.42–45.48)	0.61 (0.40–0.92)	27.27	1.10
7	6	827	1	9	40	99.9	331.20 (42.4–2584.4)	0.60 (0.40–0.91)	85.71	1.08
8	5	827	1	10	33.3	99.9	276.00 (34.3–2221.3)	0.67 (0.47–0.95)	83.33	1.19
9	3	827	1	12	20	99.9	165.6 (18.3–1501.8)	0.80 (0.62–1.03)	75.00	1.43
10	1	828	0	14	6.7	100				

Table 1; it was significantly smaller in the VL group (1.8% vs 6.6%, P<0.001). There were three failed intubations (0.3%) in the ML group, and none in the VL group.

Table 3 lists the number of true positives, true negatives, false positives and false negatives, and sensitivity, specificity, positive and negative likelihood ratios, and PTP for thresholds on the El-Ganzouri scale (EGRI), for the two groups. Reported are scores between 1 and 9, because there were no true negatives at score 0, and no true positives at score 10. The cut-off value was 2 for the ML group, and 3 for the VL group. At the cut-off value, in the ML and VL groups, respectively, sensitivity was 69.7% and 93.3%, and specificity was 66.3% and 76.6%. The corresponding positive PTP were 12.81% and 6.73%, and the corresponding negative PTP were 3.15% and 0.16%.

In the VL group, the positive PTP increased from 27.27% to 85.71% as the threshold was elevated from 6 to 7, whereas the negative PTP remained essentially unchanged at about 1.1%. In the ML group, the threshold value of 4 suggested by El-Ganzouri and colleagues[3] was associated with a positive PTP of 31.34% and a negative PTP of 4.85%.

The ROC curves for the ML and VL groups are shown in Figure 2. The AUC was 0.74 in the ML group, and 0.91 in the VL group; both values were significantly higher than 0.5 (P<0.001).

Discussion

Several risk indexes for difficult intubation have been proposed in the last decade.[1–4] Their sensitivity and specificity remain modest, and no rating system is currently considered as fail-safe.[16] As a consequence, even in centres where they are used, decisional thresholds are normally not defined. This is also the case in our institution.[16]

Studies evaluating the predictive value of risk indexes have been conducted using direct Macintosh laryngoscopy as the reference standard to determine the presence of the
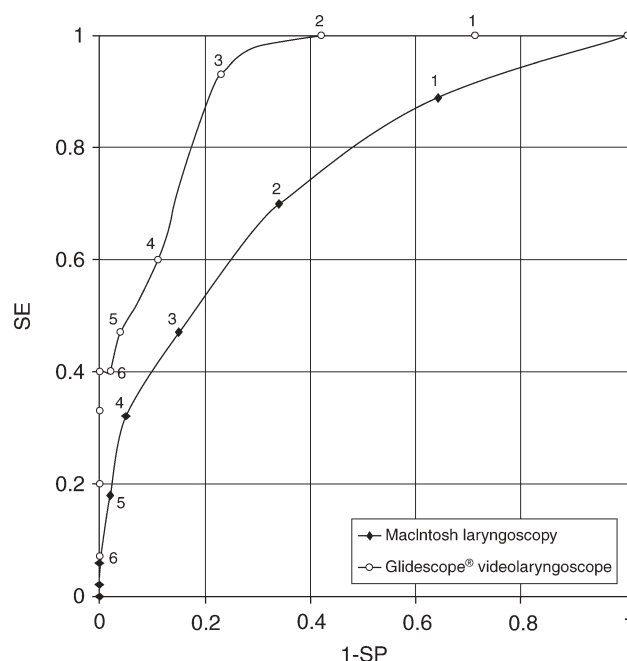
Fig 2 ROC curves for the El-Ganzouri risk index, computed using Macintosh laryngoscopy or Glidescope® videolaryngoscopy as reference standard. The ciphers along the curves are the corresponding El-Ganzouri scores. SE, sensitivity; SP, specificity.

condition of interest, namely C&L grade III–IV or grade IIa, b–III–IV.[1–4] To our knowledge, this is the first study considering Glidescope® videolaryngoscopy as reference standard.

In studies of diagnostic accuracy, the reference standard is assumed to be the best available method to establish the presence of the condition of interest; if the reference standard is not 100% accurate, the index test (i.e. the El-Ganzouri index in our case) may correctly classify cases that have been incorrectly classified by the reference standard.[17] As a consequence, the overall accuracy of the predictive indexes for difficult intubation may be underestimated if the accuracy of the reference standard is limited.[17]

Our findings suggest this may indeed be the case. In accordance with previous reports, we found that the period prevalence of high C&L grades was significantly smaller (1.8% *vs* 6.6%) with Glidescope® videolaryngoscopy compared with direct Macintosh laryngoscopy.[6] The accuracy of the El-Ganzouri index with respect to the C&L grade considerably improved, from 0.74 (considered poor according to published guidelines) to 0.91 (which is considered good).[18] By increasing the cut-off value from 2 to 3, both sensitivity and specificity increased considerably.

El-Ganzouri and colleagues[3] suggested using a decisional threshold of 4. In the ML group, this was associated with a positive PTP of 31.34% and a negative PTP of 4.85%. In the VL group, a threshold of 6 was associated with a slightly lower positive PTP (27.27%) and a considerably lower negative PTP (1.1%). In the same group, a threshold of 7 was associated with a much higher positive PTP (85.71%) and an essentially unvaried negative PTP. Such a steep variation of positive PTP is most likely a statistical artifact caused by the small number of cases with El-Ganzouri score higher than 6 (2 in the ML group and 7 in the VL group). This is also suggested by the very large confidence intervals for the positive likelihood ratio at scores higher than 6.

The differences in PTP between the two groups are partly due to improved accuracy of the index test with respect to the C&L grade, and partly due to reduced period prevalence of the unfavourable outcome.[15]

Although the observed difference in the distribution of El-Ganzouri scores between the two groups is a limitation of the present study, it may not explain the observed improvement in test accuracy and the decreased prevalence of the unfavourable outcome. Although it is possible, it is unlikely for the results of this study to be biased by the fact that they were derived from subsequently studied cohorts, as the El-Ganzouri index had been in routine use for about 2 yr before the first intubation included in the study. Also, the anaesthesiological team remained unchanged during the study period. The results of this study may, however, suffer from sources of bias related to its retrospective nature, in particular from the fact that there were multiple assessors and that the anaesthesiologist evaluating the C&L grade was not blinded to the El-Ganzouri score. Furthermore, another potential source of bias is the fact that in some cases the same anaesthesiologist who measured the index also performed laryngeal exposure and determined the C&L grade; this, however, was the case only for a relatively small proportion of cases, which was comparable for the two groups.

Although our results seem to suggest that a decisional threshold between 6 and 7 might be appropriate with Glidescope® videolaryngoscopy, this result should be interpreted with caution due to the small number of cases with high El-Ganzouri scores. Firm conclusions require further research. Also, the results of this study may not be applicable to other types of patients prone to difficult intubation, such as those undergoing obstetric or bariatric surgery.[19 20]

Conclusion

Our study suggests that the predictive value of the El-Ganzouri index may have been underestimated due to limited accuracy of Macintosh laryngoscopy as reference standard. If accuracy values on the order of those obtained in this study can be reproducibly obtained using Glidescope® videolaryngoscopy as reference, then the El-Ganzouri index may be reconsidered as a decisional tool if videolaryngoscopy is used in daily clinical practice. Further studies with larger and different patient groups would be of value in this area.

Acknowledgement

References

1 Mallampati SR, Gatt SP, Gugino LD, *et al.* A clinical sign to predict difficult tracheal intubation: a prospective study. *Can Anaesth Soc J* 1985; **32**: 429–34

2 Wilson ME, Spiegelhalter D, Robertson JA, Leser P. Predicting difficult intubation. *Br J Anaesth* 1988; **61**: 211–6

3 El-Ganzouri AR, McCarthy RJ, Tuman KJ, *et al.* Preoperative airway assessment: predictive value of a multivariate risk index. *Anesth Analg* 1996; **82**: 1197–204

4 Nath G, Sekar M. Predicting difficult intubation–A comprehensive scoring system. *Anaesth Intensive Care* 1997; **25**: 482–6

5 Glidescope® Videolaryngoscope, Verathon™ Inc. Available from http://www.verathon.com/glidescope_products.htm (last accessed on 8 August 2007)

6 Cooper RM. Use of a new videolaryngoscope (GlideScope) in the management of a difficult airway. *Can J Anaesth* 2003; **50**: 611–3

7 Rai MR, Dering A, Verghese C. The Glidescope system: a clinical assessment of performance. *Anaesthesia* 2005; **60**: 60–4

8 Sun DA, Warriner CB, Parsons DG, *et al.* The GlideScope video laryngoscope: randomized clinical trial in 200 patients. *Br J Anaesth* 2005; **94**: 381–4

9 Cuchillo JV, Rodriguez MA. Considerations aimed at facilitating the use of the new GlideScope videolaryngoscope. *Can J Anaesth* 2005; **52**: 661

10 Kramer DC, Osborn IP. More maneuvers to facilitate tracheal intubation with the GlideScope. *Can J Anaesth* 2006; **53**: 737

11 Samsoon GL, Young JR. Difficult tracheal intubation: a retrospective study. *Anaesthesia* 1987; **42**: 487–90

12 Knill RL. Difficult laryngoscopy made easy with a 'BURP'. *Can J Anaesth* 1993; **40**: 279–82

13 Frova G. Comparison of tracheal introducers. *Anaesthesia* 2005; **60**: 516–7

14 De Amici D, Klersy C. Incidence, prevalence and anaesthesia: a troubled marriage? *Eur J Anaesthesiol* 1997; **14**: 107

15 Langlotz CP. Fundamental measures of diagnostic examination performance: usefulness for clinical decision making and research. *Radiology* 2003; **228**: 3–9

16 Practice guidelines for the management of the difficult airway: an updated report by the American Society of Anesthesiologysts Task Force on management of the difficult airway. *Anesthesiology* 2003; **78**: 597–602

17 Whiting P, Rutjes A, Reitsma JB, *et al.* Sources of variation and bias in studies of diagnostic accuracy. *Ann Intern Med* 2004; **140**: 189–202

18 Jones CM, Athanasiou T. Summary receiver operating characteristic curve analysis techniques in the evaluation of diagnostic tests. *Ann Thorac Surg* 2005; **79**: 16–20

19 Voyagis GS, Kyriakis KP, Dimitriou V, Vrettou I. Value of oropharyngeal Mallampati classification in predicting difficult laryngoscopy among obese patients. *Eur J Anaesthesiol* 1998; **15**: 330–4

20 Gupta S, Pareek S, Dulara SC. Comparison of two methods for predicting difficult intubation in obstetric patients. *Middle East J Anesthesiol* 2003; **17**: 275–85