Increase in Mallampati score during pregnancy


Summary
A photographic version of the Mallampati test was developed and applied to 242 pregnant patients at 12 weeks’ gestation and again at 38 weeks’ gestation. At 38 weeks the number of grade 4 cases had increased by 34% (P < 0.001). This is in agreement with other evidence which suggests that difficult laryngoscopy is slightly more frequent in obstetrics (1.7%) than in general surgery (1.3%). The increase in Mallampati score correlated with gain in body weight (r = 0.3, P < 0.001), which gives some support to the concept that fluid retention is the underlying cause. We conclude that pharyngeal oedema causes some hindrance to tracheal intubation in obstetrics, but not enough to explain the high failure rate reported. A case is made for rationalizing the management of difficult intubation. Our data also show that more research is needed on factors which affect Mallampati’s test, particularly neck extension. (Br. J. Anaesth. 1995; 74: 638-642)

Key words

There has so far been no satisfactory explanation for the high incidence of failed intubation in obstetrics, eight times larger than in general surgery [1]. Factors such as enlarged breasts may be implicated, but simple manoeuvres for dealing with these problems were described many years ago. One factor that we discussed previously [2], pharyngeal oedema, is harder to assess and is the subject of the present study. We argued that if pharyngeal oedema were a significant factor then the Mallampati score should increase during pregnancy. However, before testing this it seemed advisable to try to improve the reproducibility of the method.

There are large discrepancies in the results reported with Mallampati’s method; for example, in the original study [3] the highest grade occurred in 15 of 210 patients (7%), in contrast with 114 of 675 (17%) in another study [4]. This discrepancy is too large to be fortuitous (P = 7 x 10^-4) and it seems likely that the test was performed differently by the two teams. We decided therefore to standardize the method as much as possible.

Patients and methods
With Ethics Committee approval, all patients booked for the antenatal clinic were sent a letter describing the study and, on attendance, consent was obtained. Patients with significant neck pathology were excluded. The study was carried out early and late in pregnancy in 261 patients, but only those with complete data sets were analysed; 242 in the Mallampati study, 230 in the grade/weight study.

POSITION OF THE HEAD
The original study [3] did not specify the position of the patient’s head. We opted for the neutral position; the tip of the ear was aligned with the outer canthus of the eye, using a horizontal indicator (fig. 1).

POSITION OF THE CAMERA
Hitherto the position of the observer’s eye relative to the patient was inevitably variable; to eliminate this source of error, a camera was fixed at the level of the patient’s pharynx and 30 cm from it. Focal length was 90 mm, through-the-lens metering, ring flash and automated colour printing were used throughout.

To assess observer error, the photographs were evaluated independently by three anaesthetists (consultant, senior registrar and registrar); to assess patient error, two photographs were taken of each patient at each period. Both photographer and patients were told that the appearance of the throat during pregnancy was to be studied, but not the changes expected. The observers may have guessed the object of the study, but had no way of knowing if a photograph was from period 1 or 2. The photographs were coded, issued in random order and decoded at the end of the study. The data were entered in a spreadsheet from which they could be exported to other programs for analysis, using a standard “386” machine with maths co-processor.

S. PILKINGTON*, MB, BS, FRCA, F. CARLI†, DMS, MPhil, FRCA, M. J. DAKIN, MB, BS, FRCA, M. ROMNEY, SRN, K. A. DE WITT, BIPP, C. J. DORÉ‡, BSc, R. S. CORMACK, BA, BSc, BM, BCH, FRCA, Divisions of Anaesthesia, Medical Photography and Medical Statistics, Northwick Park Hospital and Clinical Research Centre, Harrow HA1 3UJ. Accepted for publication: December 24, 1994. Present addresses: *Department of Anaesthesia, University of Natal, Durban, South Africa. †McGill University, Royal Victoria Hospital, Montreal H3A 1A1, Canada. ‡Medical Statistics Unit, Royal Postgraduate Medical School, Hammersmith Hospital, DuCane Road, London W12 0NN. Correspondence to F.C.
Mallampati score in pregnancy

Figure 1 Camera set to obtain photographs for later evaluation of the Mallampati grade. Patient’s head in neutral position. Heights and distances between camera and patient are fixed.

STATISTICAL METHODS

Wilcoxon’s signed rank test was applied to the change in Mallampati grades, using paired data from 12 and 38 weeks’ gestation on each patient. Reproducibility was estimated by the kappa test, a chance-corrected measure of agreement in which 1 implies complete agreement. It is described in detail by Fleiss [5]. The grade/weight correlation raised some problems. If fluid retention is the cause of the increase in Mallampati score then some relationship would be expected between gain in body weight and gain in grade. The regression line should go through the origin; routine least squares analysis confirmed that the intercept was not significantly different from zero (P = 0.4). Next a least squares regression line through the origin was undertaken [6]. However, significance testing was complicated because the residuals were very far from Gaussian. The residuals are the deviations of each point from the line of best fit. To test for normality, each residual was plotted against its rankit (a rankit resembles a standardized normal deviate, with various corrections). This is known as a “normal plot of residuals” (fig. 3) and it is a straight line if the data are perfectly Gaussian. The amount of deviation from a straight line provides an objective measure of departure from normality; this is the basis of the W’ test [7] which is more powerful than the traditional tests for skewness and kurtosis.

SIGNIFICANCE TESTING WITH NON-GAUSSIAN RESIDUALS

In the past this problem was handled by using non-parametric methods, such as Kendall’s rank correlation, or by transforming the data. However, the modern computer provides a new option, with advantages that we have detailed elsewhere [8]. The P value is determined, not by theoretical methods, but by experiment, using a program written specifically for the problem. A random number generator is used to create data with the same distribution as the real data and the mean slope is adjusted to zero in order to test the null hypothesis of no correlation. The process is repeated many times and the “experimental P” taken as the proportion of cases in which the simulated slope is as large as, or larger than the observed slope, in each direction.

DEFINITIONS

The Cormack–Lehane [9] classification of laryngoscopy and Mallampati’s test [3] both define four different “grades”. To avoid confusion these are referred to as laryngoscopy grades, or Mallampati grades, as appropriate. Originally there were three Mallampati grades; a fourth was added by Samsoon and Young [10] and we used that system: grade 1 = faucial pillars, soft palate and uvula all visible; grade 2 = faucial pillars and soft palate visible, but not the uvula; grade 3 = only soft palate visible; grade 4 = not even soft palate visible.

Caveats against phonation [11] and for ensuring patient co-operation [12] were observed.

Results

Height was found to follow a Gaussian distribution and so the mean (SD) is shown. All other data were highly skewed, to the left or right, and for these the median (95% range) are shown: mean height 162 (SD 6.7) cm; median gestation, 1st period: 12 (95% range 8–14) weeks; median gestation, 2nd period: 38 (34–40) weeks; weight, 1st period: 60.3 (41.8–79.2) kg; weight, 2nd period: 71.3 (53.0–94.5) kg.

MALLAMPATI GRADES AT 12 AND 38 WEEKS

Each of the three observers carried out 968 Mallampati tests, as summarized in table 1 and figure 2. Duplicate photographs of each patient were labelled photos (1) and (2) at 12 weeks’ gestation, (3) and (4) at 38 weeks’ gestation. With three observers there are 12 possible comparisons between the two periods. All 12 comparisons showed a similar picture, that is

<table>
<thead>
<tr>
<th>Photos</th>
<th>(1)</th>
<th>Mean</th>
<th>(3)</th>
<th>(4)</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grade 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Grade 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Grade 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Grade 4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 1 Distribution of Mallampati grades at 12 and 38 weeks’ gestation. Each pair of duplicate photographs were taken: photos (1) and (2) at 12 weeks’ gestation, photos (3) and (4) at 38 weeks’ gestation. Each photograph was assessed independently by three different observers.

DEFINITIONS

The Cormack–Lehane [9] classification of laryngoscopy and Mallampati’s test [3] both define four different “grades”. To avoid confusion these are referred to as laryngoscopy grades, or Mallampati grades, as appropriate. Originally there were three Mallampati grades; a fourth was added by Samsoon and Young [10] and we used that system: grade 1 = faucial pillars, soft palate and uvula all visible; grade 2 = faucial pillars and soft palate visible, but not the uvula; grade 3 = only soft palate visible; grade 4 = not even soft palate visible.

Caveats against phonation [11] and for ensuring patient co-operation [12] were observed.

Results

Height was found to follow a Gaussian distribution and so the mean (SD) is shown. All other data were highly skewed, to the left or right, and for these the median (95% range) are shown: mean height 162 (SD 6.7) cm; median gestation, 1st period: 12 (95% range 8–14) weeks; median gestation, 2nd period: 38 (34–40) weeks; weight, 1st period: 60.3 (41.8–79.2) kg; weight, 2nd period: 71.3 (53.0–94.5) kg.

MALLAMPATI GRADES AT 12 AND 38 WEEKS

Each of the three observers carried out 968 Mallampati tests, as summarized in table 1 and figure 2. Duplicate photographs of each patient were labelled photos (1) and (2) at 12 weeks’ gestation, (3) and (4) at 38 weeks’ gestation. With three observers there are 12 possible comparisons between the two periods. All 12 comparisons showed a similar picture, that is
As we have seen (table 2), any observer, assessing any photograph at each period, would have proved the point, but that would shed no light on the errors. The percentage of grade 4 cases is seen to increase at 38 weeks with corresponding reductions in the other three grades.

SIGNIFICANCE TESTING

When multiple comparisons are made, some differences may appear significant purely by chance. Procedures such as Bonferroni's correction are designed to deal with these problems, but in this case a straightforward common sense argument can be applied. If it is a chance difference then pooling the data makes $P$ larger; if it is real then the opposite applies. Table 2 shows the latter to be true; progressive pooling of the data caused a steady reduction in the $P$ values, reinforcing the conclusion that this is a real difference.

REPRODUCIBILITY OF THE MALLAMPATI SCORE

As we have seen (table 2), any observer, assessing any photograph at each period, would have proved the point, but that would shed no light on the errors. The percentage of grade 4 cases is seen to increase at 38 weeks with corresponding reductions in the other three grades.

Table 3 “Kappa” values, which assess reproducibility of a set of measurements; high values mean good reproducibility. Here it tests agreement between: two observers assessing the same photograph; and two photographs taken on the same patient assessed by the same observer.

<table>
<thead>
<tr>
<th>Photos</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observer 1 vs 2</td>
<td>0.85</td>
<td>0.82</td>
<td>0.85</td>
<td>0.85</td>
<td>0.84</td>
</tr>
<tr>
<td>Observer 1 vs 3</td>
<td>0.78</td>
<td>0.71</td>
<td>0.78</td>
<td>0.78</td>
<td>0.76</td>
</tr>
<tr>
<td>Observer 2 vs 3</td>
<td>0.75</td>
<td>0.74</td>
<td>0.75</td>
<td>0.78</td>
<td>0.75</td>
</tr>
<tr>
<td>Grand mean</td>
<td>0.78</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Observers</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Agreement</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>between</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>observers</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Observer 1 vs 2</td>
<td>0.59</td>
<td>0.64</td>
<td>0.57</td>
<td>0.60</td>
</tr>
<tr>
<td>Observer 3 vs 4</td>
<td>0.65</td>
<td>0.70</td>
<td>0.67</td>
<td>0.67</td>
</tr>
<tr>
<td>Grand mean</td>
<td>0.64</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

By duplicating the photographs, triplicating the observers and applying the kappa test [5], the reproducibility of the method is revealed (table 3).

Agreement between observers was good (0.79), but agreement between duplicate photographs was less good (0.64). Kappa can be calculated for each grade separately, for example, by finding kappa for grade 4 against an amalgam of the other three grades. Kappa for grade 4 showed the best reproducibility (0.84), grade 1 was next (0.71), while the middle grades were worst (0.54 and 0.66). The meaning of this is analysed in more detail in a separate study.

GRADE/WEIGHT CORRELATION

The gain in Mallampati grade was plotted against gain in weight and the slope was found to be 0.02 grade kg$^{-1}$, using the method of least squares. This assumes that the residuals are Gaussian, so this was analysed next (fig. 3). On the left the normal plot of the residuals deviates widely from a straight line, showing the characteristic features of leptokurtosis [7] (too much probability in the tails). Significance tests on the residuals gave these results: skewness $P = 0.17$, kurtosis $P = 0.005$, while for the $W'$ test $P = 2 \times 10^{-4}$. In contrast, on the right in figure 3 is a typical normal distribution, the patients' heights.

The next step was to find out if the observed slope could have arisen by chance, using a method which does not assume Gaussian residuals. Figure 4 shows the results of 30000 simulations; 99% of the slopes were less than 0.01, so the observed value of 0.02 was unlikely to be caused by chance. However, although the correlation was real it was also rather weak ($r = 0.3$), because the residual SD was larger than the effect measured.

Discussion

RESEARCH INTO FAILED INTUBATION

Over the past decade a consensus has emerged internationally that the most effective first method of handling difficult intubation is to use a flexible
introducer (gum elastic or soft wire). This is effective for all laryngoscopy grade 2 cases, most grade 3 cases, but not grade 4. It follows that the incidence of grades 3 and 4 provides an objective measure of the risk of failed intubation. Grade 4 is very rare and so, for practical purposes, laryngoscopy grade 3 frequency is relevant. However, one caveat is needed; our recent study [2] showed that if laryngoscopy is not optimal then false high values for grade 3 incidence are obtained.

When care was taken to use optimal laryngoscopy [13] the incidence of laryngoscopy grade 3 cases in general surgery was 10 of 778 (1.3%). If pregnancy causes a 34% increase in Mallampati grade 4 then one might reasonably expect the laryngoscopy grade 3 incidence to increase by a similar amount, that is from 1.3% in general surgery to 1.7% in Caesarean section. This fits well with the Durban survey [14] in which 26 laryngoscopy grade 3 cases occurred in 1500 obstetric anaesthetics (1.7%). It also fits with our obstetric audit [15] with 16 grade 3 cases in 928 consecutive intubations (1.6%).

Can this rather small increase in laryngoscopy grade 3 incidence account for the eight-fold increase in frequency of failed intubation that has been reported? It seems likely that there are other factors. One that does not seem to have been considered is the policy of early abandonment of intubation. In some maternity units junior staff are advised to abandon intubation if the first attempt fails; this must increase the failure rate and it seems advisable to re-examine the policy.

Repeated attempts at intubation are contra-indicated in two situations, ineffective cricoid pressure and progressive hypoxia. The logical first step is to check that the pharynx is dry. If cricoid pressure is being applied correctly, but is not working, then not even one attempt at intubation should be made. The patient should be put head down and the failed intubation drill started. Similarly, if hypoxia is developing and the patient’s lungs cannot be ventilated while maintaining cricoid pressure, attention must focus on restoring oxygen supply. This may require trans-tracheal ventilation. It has been accepted policy since 1980 [16] to have the equipment needed for this in all maternity theatres.

However, both eventualities are extremely rare. In the majority of cases cricoid pressure prevents regurgitation and the patient’s lungs may be ventilated, if necessary, while maintaining cricoid pressure. In that event there is no desperate hurry; several intubation attempts can safely be made. The “give up early” policy originated from the concept of “crash induction” in which speed was held to be the main safety factor, but this is now widely seen as a mistake; the most important safety factor is not speed, but cricoid pressure. Our original advice [9] to make only three attempts at intubation was based, not on the need for speed, but on our experience that if three attempts fail then further efforts are unlikely to succeed. In a few grade 3 cases the larynx is displaced too far forward for easy access by the introducer. In short, if the guidelines continue to evolve in a rational way then failed intubation should become nearly as rare in obstetrics as in general surgery.
PREDICTING DIFFICULT INTUBATION

Effective prediction requires a combination of tests, which is likely to include Mallampati's, so it seems worth trying to perfect it. In this study, the incidence of Mallampati grade 4 cases at the end of pregnancy (56%) was much higher than in the Durban survey (5%). Ethnic differences might account for this, but it seems unlikely because the frequency of difficult laryngoscopy is very similar in the two populations (1.7%, 1.6%). A more plausible explanation is that we opted to put the head in the neutral position. Extending the head on the atlas (as in the Magill position) is likely to improve the view and reduce the Mallampati score. However, we chose the neutral position because our immediate aim was to detect a small change in Mallampati score, so it was essential to maximize reproducibility. Ability to extend the neck varies; a constant neutral position seemed preferable. On the other hand, when the aim is to predict difficult intubation the neutral position is probably not the best, because it gives 56% of Mallampati grade 4 cases when only 1.7% are at risk of failed intubation.

Whatever position is chosen it seems advisable to record it. One other team, the Cardiff group [12], specified the neutral position, but did not fix it. An observer who stands beside the patient while the test is being carried out may note that, even when ostensibly neutral, the position is more extended than in our figure 1. The Cardiff team also commented that their Mallampati score was much higher than in the original study. We conclude that the head position deserves further study as a possible explanation for the discrepancies in the literature.

If there is some moment to moment variation in the view displayed by the patient (as our photographs indicate), then with the test as normally carried out the observer is likely to select the best view. On the other hand, a photograph freezes just one frame, as it were, not necessarily the best, and this would tend to increase the Mallampati score.

The Cardiff team also reported that their two observers sometimes gave conflicting verdicts on the same patient, adding that it was not possible to say if this was caused by observer error or to a change in the view presented by the patient. Our photographic method solves that problem; clearly different verdicts on the same photograph can only arise from observer error. From this it emerges, rather surprisingly, that the variation arises mainly from the patient (despite our efforts to standardize the test), not from the observer. This is analysed more precisely in a separate study designed to disentangle and measure the various sources of error.

The correlation between weight gain and increase in Mallampati score lends some support to the suggestion that fluid retention causing pharyngeal oedema is the underlying process. However, the correlation is too weak to be clinically useful, which underlines the need to reduce the grade standard deviation. It should help to have an index of fluid retention which is more specific than weight; if such an index showed a strong correlation with grade gain then it might be useful as part of a system for predicting difficult intubation.

Acknowledgements

We thank Dr V. Ramachandra, Dr R. Aquilina and Dr M. Shaw, who each assessed over 1000 photographs.

References