Endoscopic assessment of oesophagitis: clinical and functional correlates and further validation of the Los Angeles classification


Abstract

Background—Endoscopic oesophageal changes are diagnostically helpful and identify patients exposed to the risk of disease chronicity. However, there is a serious lack of agreement about how to describe and classify the appearance of reflux oesophagitis.

Aims—To examine the reliability of criteria that describe the circumferential extent of mucosal breaks and to evaluate the functional and clinical correlates of patients with reflux disease whose oesophagitis was graded according to the Los Angeles system.

Methods—Forty six endoscopists from different countries used a detailed worksheet to evaluate endoscopic video recordings from 22 patients with the full range of severity of reflux oesophagitis. In separate studies, Los Angeles system gradings were correlated with 24 hour oesophageal pH monitoring (178 patients), and with clinical trials of omeprazole treatment (277 patients).

Results—Evaluation of circumferential extent of oesophagitis by the criterion of whether mucosal breaks extended between the tops of mucosal folds, gave acceptable agreement (mean $\kappa$ value 0.4) among observers. This approach is used in the Los Angeles system. An alternative approach of grouping the circumferential extent of mucosal breaks as occupying 0–25%, 26–50%, 51–75%, 76–99%, or 100% interspersed with the Los Angeles classification.

Severity of oesophageal acid exposure was significantly ($p<0.001$) related to the severity grade of oesophagitis. Pretreatment oesophagitis grades A–C were related to heartburn severity ($p<0.01$), outcomes of omeprazole (10 mg daily) treatment ($p<0.01$), and the risk for symptom relapse off therapy over six months ($p<0.05$).

Conclusions—Results add further support to previous studies for the clinical utility of the Los Angeles grading system for endoscopic grading of oesophagitis.

(Gut 1999;45:172–180)

Keywords: oesophagitis; endoscopy; stricture; columnar lined mucosa; heartburn; omeprazole; acid reflux
Table 1  The Los Angeles classification of oesophagitis

<table>
<thead>
<tr>
<th>Grade</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grade A</td>
<td>One (or more) mucosal break no longer than 5 mm, that does not extend between the tops of two mucosal folds</td>
</tr>
<tr>
<td>Grade B</td>
<td>One (or more) mucosal break more than 5 mm long that does not extend between the tops of two mucosal folds</td>
</tr>
<tr>
<td>Grade C</td>
<td>One (or more) mucosal break that is continuous between the tops of two or more mucosal folds but which involves less than 75% of the circumference</td>
</tr>
<tr>
<td>Grade D</td>
<td>One (or more) mucosal break which involves at least 75% of the oesophageal circumference</td>
</tr>
</tbody>
</table>

We analysed the data from the large number of observers as a multiple of comparisons between each pair of observers. That is, if 46 observers recorded the presence or absence of a particular feature, κ statistics were calculated for a total of \( \binom{n}{2} \) comparisons. This means that a theoretical number of 2070 κ values have been computed for each item. Each κ value measured the agreement between two observers in the endoscopic videotape recordings in the 22 patients; κ values are given as medians and interquartile ranges.

We used a detailed worksheet (shown in full in the Appendix), which was a development of the one used in previous studies. The sheet scored the full range of findings relevant to the evaluation of an oesophagitis classification system, according to predefined criteria. The previously agreed definition of mucosal breaks, “an area of slough or erythema with a discrete line of demarcation from the adjacent more normal looking mucosa” was used. The original videotapes were acquired with Olympus video endoscopes and either a Super-VHS or NT video recorder. The copies that were distributed were made directly from these original recordings. Five centres participated in producing original video material.

The most widely used coefficient of agreement in clinical studies, the κ statistic, was used in its original version in order to evaluate the degree of agreement among observers. The range of possible values for κ is from −1, which indicates perfect disagreement, to +1, which is reached with perfect agreement. Agreement by chance alone gives a value of 0. Though far from perfect, values as low as 0.4 are considered to indicate acceptable agreement. In the κ statistics used, p0 is the observed proportion of agreement and pC is the expected agreement by chance in the relevant contingency table.

\[ \kappa = \frac{p_0 - p_C}{1 - p_C} \]

We analysed the data from the large number of observers as a multiple of comparisons between each pair of observers. That is, if 46 observers recorded the presence or absence of a particular feature, κ statistics were calculated for a total of \( \binom{n}{2} \) comparisons. This means that a theoretical number of 2070 κ values have been computed for each item. Each κ value measured the agreement between two observers in the endoscopic videotape recordings in the 22 patients; κ values are given as medians and interquartile ranges.
six months, or if their major symptom was acid regurgitation, abdominal pain, or discomfort in association with heartburn. In total, 178 patients were entered into the study from 17 centres in Denmark, Norway, and Sweden. The patients scored their reflux induced symptoms in a structured manner during the last two days before enrolment into the study as either mild, moderate, or severe. At endoscopy, oesophagitis was graded according to the Los Angeles classification. Before entry into the therapeutic trial ambulatory 24 hour pH monitoring was carried out with a monocrystalline, antimony electrode positioned 5 cm above the oral margin of the lower oesophageal sphincter, which was located by stationary oesophageal manometry. During the pH study, the patients were asked to act as normally as possible in order to minimise any impact on reflux patterns that could result from the process of measurement.

Study II—Clinical correlates
In the other clinical study, the major therapeutic outcomes of which have been reported elsewhere,11 patients with upper gastrointestinal symptoms were recruited from the primary care setting. The patients were screened for reflux symptoms by use of a questionnaire,12 which had been developed for the diagnosis of symptomatic reflux disease, and which focuses especially on objective recognition of heartburn, and on exacerbating or relieving factors. Patients with a questionnaire score of 4 or more, aged 18–80 years, were screened by endoscopy for enrolment into the study, provided that they had been experiencing upper gastrointestinal symptoms for at least three months with episodes occurring on at least four days during the previous week. At endoscopy, oesophagitis was graded according to the Los Angeles classification (table 1). For ethical reasons, patients found to have Los Angeles grade D oesophagitis, columnar lined oesophagus, or peptic strictures were excluded from further involvement in the study. Complete endoscopic and clinical data were available for 496 of the 538 patients randomised to treatment (see below). Of these, the 277 who had mucosal breaks were randomised to either omeprazole 20 or 10 mg daily. Patients without mucosal breaks were randomised to either omeprazole 20 or 10 mg daily or to placebo. Treatment was given for four weeks, under double blind conditions, after which it was stopped in those who had complete symptom relief. Those with persistent symptoms at four weeks were given open therapy with omeprazole 20 mg daily for another four weeks. All patients in whom symptoms were relieved at four or eight weeks entered a follow up period without treatment for up to six months, exiting the study if their symptoms relapsed according to predefined criteria before that time. There were 145 patients who had mucosal breaks at their pretreatment endoscopy who entered this phase of the study.

STATISTICS
The relation between acid reflux variables and endoscopic grading of mucosal breaks was assessed with a logistic regression analysis in which the endoscopic grade was the dependent variable. The relation between symptom severity and endoscopic grading was evaluated by cross tabulation of the severity of heartburn and endoscopic grading and then a simple correlation analysis. A log rank test was applied to the proportion of patients in clinical remission during follow up with the pretreatment endoscopic grading as the dependent variable.

Results
INTEROBSERVER AGREEMENT
There was no significant difference in the levels of agreement on judgements described below between experienced endoscopists and trainees, so the entire data set was pooled. The most important aspect of the endoscopic video sequence assessments was the reproducibility of evaluation of the circumferential extent of mucosal breaks as judged by the extension of mucosal breaks across the tops of two or more mucosal folds. Such mucosal breaks were assessed with the $k$ value of 0.4 (0.22–0.51) (fig 1).

When endoscopists were asked to evaluate the proportion of the circumference involved by mucosal breaks as 0–25%, 26–50%, 51–75%, 76–99%, or 100%, only the most limited extent (0–25%) was recorded with any level of accuracy ($k = 0.4$, range 0.22–0.52). In the remaining groups, $k$ values equalled or were close to 0. Another aspect of assessing the circumferential extent was the assessment of whether one or more mucosal folds were involved by the mucosal breaks. The examiners again showed an acceptable agreement, scoring a $k$ value of 0.4 in this judgement.

The score sheets also allowed assessment of interobserver variation in the recognition of other aspects of the endoscopic oesophageal appearances already evaluated in our previous study. Agreement on the presence or absence of a mucosal break in a particular endoscopic video sequence had a $k$ value of 0.4 (0.21–0.58). The presence of multiple mucosal breaks (two or more) was assessed with a similar level of accuracy ($k$ value 0.4 (0.25–0.54)).
The presence of a stricture was assessed with a $\kappa$ value of only 0.33 (0–0.46) but there were only two cases of stricture among the 22 cases. For the four cases of columnar metaplasia (all histologically confined), recognition of tongues that extended up the oesophagus had a $\kappa$ value of 0.38 (0.13–0.58) and estimates of the circumferential extent of this metaplastic epithelium had a $\kappa$ value of 0.42 (0.10–0.65). The investigators were specifically asked to assess the number and depth of ulcers, as opposed to mucosal breaks, with an ulcer being defined as a mucosal break greater than 3 mm deep. Reliable judgement on the presence of an ulcer was not achieved as this assessment had a $\kappa$ value of −0.01 and the corresponding $\kappa$ value for the assessment of the number of ulcers only reached −0.02.

The presence and absence of “minimal endoscopic changes” as listed in table 2 was evaluated close to the oesophagogastric mucosal junction; there was no agreement on the presence or absence of these findings.

The presence and absence of “minimal endoscopic changes” as listed in table 2 was evaluated close to the oesophagogastric mucosal junction; there was no agreement on the presence or absence of these findings.

### Table 2  $\kappa$ values for the assessments of minimal changes in the distal oesophagus and the gastro-oesophageal junction

<table>
<thead>
<tr>
<th>Endoscopic feature</th>
<th>$\kappa$ value</th>
<th>Median $\kappa$ value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Excessive reddening of the cardia</td>
<td>0.09</td>
<td>(0.00–0.29)</td>
</tr>
<tr>
<td>Erythema of the squamocolumnar junction</td>
<td>0</td>
<td>(−0.08–0.06)</td>
</tr>
<tr>
<td>Friability of the squamocolumnar junction</td>
<td>0.04</td>
<td>(0.00–0.25)</td>
</tr>
<tr>
<td>Blurring of the squamocolumnar junction</td>
<td>0.08</td>
<td>(0.00–0.22)</td>
</tr>
<tr>
<td>Diffuse erythema of the distal oesophagus</td>
<td>0.08</td>
<td>(0.00–0.28)</td>
</tr>
<tr>
<td>Patchy erythema of the distal oesophagus</td>
<td>0.00</td>
<td>(−0.07–0.09)</td>
</tr>
<tr>
<td>Increased vascularity of the distal oesophagus</td>
<td>0.00</td>
<td>(0.00–0.09)</td>
</tr>
<tr>
<td>Oedema or accentuation of the mucosal folds</td>
<td>0.00</td>
<td>(−0.02–0.15)</td>
</tr>
</tbody>
</table>

Median $\kappa$ values are given, as well as the first and third quartile.

### Table 3  Acid reflux parameters in relation to endoscopic grading of mucosal breaks

<table>
<thead>
<tr>
<th>Grade</th>
<th>Number of patients</th>
<th>% time pH &lt;4</th>
<th>Number of reflux episodes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal</td>
<td>40</td>
<td>6.7 (1.0)</td>
<td>59.5 (12.8)</td>
</tr>
<tr>
<td>A</td>
<td>50</td>
<td>9.3 (1.3)</td>
<td>106.9 (12.8)</td>
</tr>
<tr>
<td>B</td>
<td>50</td>
<td>13.7 (2.1)</td>
<td>138.3 (20.0)</td>
</tr>
<tr>
<td>C</td>
<td>9</td>
<td>11.7 (3.0)</td>
<td>135.8 (29.4)</td>
</tr>
<tr>
<td>D</td>
<td>10</td>
<td>19.1 (6.6)</td>
<td>163.0 (41.7)</td>
</tr>
<tr>
<td>C + D</td>
<td>19</td>
<td>15.6 (3.8)</td>
<td>150.1 (25.5)</td>
</tr>
</tbody>
</table>

Results expressed as mean (SEM). Normal versus grade A, B, C, D, $p=0.0007$; normal + grade A versus grade B, C, D, $p=0.003$; normal + A, B versus grade C, D, $p=0.01$.

**Figure 3**  The relation between pretreatment endoscopic grading of mucosal breaks and healing of these mucosal breaks after four weeks of omeprazole 10 or 20 mg daily.

**Figure 4**  Percentage of patients in clinical remission during six months, follow up after initial treatment with omeprazole, which had healed the oesophagitis and controlled symptoms. The symptom relapse curves are given for each pretreatment endoscopic grade. Normal versus grade A, $p=0.04$; normal versus grade A + B + C, $p=0.002$; normal + grade A + B versus grade C, $p=0.003$.

**FUNCTIONAL AND CLINICAL CORRELATES**

Oesophageal pH monitoring and symptom assessments were completed in 178 patients as a prelude to inclusion in the short term therapeutic trial with omeprazole (table 3). There was a significant relation between the Los Angeles classification grade and the 24 hour oesophageal acid exposure values ($p<0.01$). There was also a significant correlation between pretreatment symptom severity and endoscopic grading of mucosal breaks (fig 2).
The severity of heartburn correlated significantly with the pretreatment endoscopic grade in both studies (study I: \( r=0.31, p<0.001 \); study II: \( r=0.23, p<0.01 \)).

After four weeks of treatment with omeprazole 10 mg daily, there was a gradation of endoscopic healing rates from 77% of grade A patients to only 20% in grade C patients (fig 3). There was a gradation of healing rates with omeprazole 20 mg, but this time only between grades B and C.

Symptomatic relapse after initial short term treatment occurred in 83% of the patients during a six month follow up period. The proportion of patients still in clinical remission during these six months showed a significant relation to the pretreatment Los Angeles classification grade of mucosal breaks (fig 4).

Discussion
This report provides data which support the approaches used by the Los Angeles classification to categorise the extent of oesophagitis. Of special interest was the evaluation of the circumferential extent of the mucosal breaks. The data were generated by a newly structured worksheet and new endoscopic images better directed to the aim of the assessment than the one used previously. These results indicate that endoscopists can score the joining of mucosal breaks between the tops of two mucosal folds with fair agreement, thus allowing scoring of circumferential extent with this method. Determination of radial extent without reference to mucosal folds had very poor levels of agreement. This report gives clinical and physiological data, which support the practical relevance of the Los Angeles classification and also provide the definitive description of the finally agreed form of the Los Angeles classification.

The version of the Los Angeles system described in the previous report of this group was a proposed system still under evaluation, which has now been modified slightly from the original description (table 1). If, in the future, further data suggest that the Los Angeles classification could be improved by some evolution of the criteria, it is important that even a minor modification should be given an entirely different name, in order to avoid any ambiguity as to what criteria have been used. This is an important lesson, which has not been learnt from the Savary and Miller grading system. The first version of this grading system differs substantially from the second, yet it is rarely stated which version is being used, either in clinical practice or even in published clinical trials. If the version in use is stated, the reader must be familiar with the detail of each version.

The major aim of the assessments of the endoscopic images presented in this report was to evaluate further how reliably endoscopists could assess the circumferential extent of mucosal breaks by localising the peaks of the mucosal folds. It was believed that the findings of the assessment from our previous study could have been influenced by technical limitations of the assessment method. The present study gives a more positive indication that the peaks of the oesophageal mucosal folds are useful endoscopic landmarks for determination of extent. Another important message, related to the clinical usefulness of the system, was that no apparent differences emerged between experienced endoscopists and trainees. We would have liked to have obtained a stronger \( r \) value than 0.4 to support our recommendation for use of the peaks of the mucosal folds as a primary method for determining radial extent. Determination of radial extent is, however, a necessary judgment for endoscopic classification of oesophagitis, and the peaks of the mucosal folds appear to be the best option on the basis of the data obtained.

The quality of the stored images continued to be a significant technical limitation. The stored images did not fully emulate the sensitivity of a live endoscopy done by an observer orientated to the use of the peaks of the mucosal folds as radial landmarks. Furthermore, despite our best efforts, the resolution of some of the endoscopic images was not optimal. Image quality was also impaired by the processes of image storage and copying. The stored images sent out for evaluation did not allow the observer to “explore” the oesophagus in order to clarify appearances. Such clarification should always include partial deflation of the oesophagus to recognise the position of the folds. Similar limitations with the current methodology might explain the differences in \( r \) values between the present investigation and that performed by Bytzer et al. However, significant advances in the optical resolution of video endoscopes, and in the capture, storage, and reproduction of video endoscopic images in digital form are currently being introduced. We believe that these technical developments will largely overcome the problems caused by suboptimal resolution of images as currently used. Our working group plans to use these developments for preparation of educational material, but judged that it was inappropriate to delay formal introduction and evaluation of the Los Angeles classification system pending the results of a further assessment done with technically better images. It needs, however, to be emphasised that the intraobserver variability in these situations also requires further validation.

Previously, no attempt has been made to define the endoscopic appearances that should be taken as indicative of erosion or ulceration. We have described these collectively as “endoscopic mucosal breaks” for reasons given previously. The definition, “an area of slough or erythema with a discrete line of demarcation from the adjacent, more normal looking mucosa”, still requires interpretation by the observer, but it is at least more specific than no definition at all, which has been the case until now. Currently, some members of the working group are conducting a study which correlates endoscopic mucosal breaks with the histology of biopsies directed precisely to sample these areas. The group will also consider other measures that may aid the distinction of...
Endoscopic assessment

177

resolution video endoscopy and dye spraying hold some promise for this include high
patches or tongues of mucosal columnar
may make these changes more reliable in the future. Until “minimal changes” are shown to be of
diagnostic value, they probably should not be
to realise the pivotal role
are obtained for recognition of “minimal changes”, in keeping with results from another
group.7 Again we have to consider that current evidence indicates that reliable recognition of “minimal changes” is problematical and that the diagnostic significance of these changes for reflux disease is uncertain. Possibly, better image quality may
are shown to be of
diagnostic value, they probably should not be
recorded as present or absent, as to do so only causes confusion.

The Los Angeles classification was developed with the intention to provide a clinically relevant stratification of the severity of oesophagitis. Despite the lack of validation of other oesophagitis classification systems there is considerable data indicating that these recognise clinically relevant variations of the severity of oesophagitis. This is the case for both oesophageal pH monitoring data and response to medical treatment.19–21 The oesophageal acid exposure data included in the present study support the clinical relevance of the Los Angeles classification with the possible exception of the acid reflux values in grade C patients. However, data were available from only nine patients with this oesophagitis grade, a number that is probably too small to draw a meaningful conclusion, given the obvious variability of pH monitoring acid exposure data.

Large prospective, randomised clinical trials gave us the opportunity to evaluate further the clinical relevance of the Los Angeles classification. Previous studies, which have used different endoscopic classification systems to correlate the severity of symptoms with endoscopic grade, have given negative results.7 35 In the present study there was a significant correlation between the severity of heartburn and the Los Angeles grading. The clinical studies also found an impressive gradation of endoscopic healing relating to the endoscopic grading, when omeprazole was given in a dose of 10 mg daily for four weeks. The large numbers of patients involved and the statistically significant differences are arguments for the clinical relevance of the Los Angeles classification. On the other hand, the size of the values (r=0.23–0.31) emphasises that factors other than endoscopic grading play a significant role in the clinical manifestation of the disease. It is worth noting that the data obtained show that the classification of patients into grade A and B seems to provide a clinically relevant subdivision. These patients represent the most numerous groups of patients suffering from reflux diseases, which have previously been grouped somewhat ambiguously into one grade by the Savary and Miller classification.36 Additional information relevant to the significance of the Los Angeles classification system was gained from following patients after a successful initial, short term, drug therapy. The symptomatic relapse curves separated patients, who before treatment had an absence of oesophageal mucosal breaks, from those with Los Angeles grades A–C.

The findings presented in this and our previous report,7 represent the only rigorous assessment of an endoscopic grading system for reflux oesophagitis. The criteria used for the Los Angeles classification have been developed and agreed on by a truly international working group. It is hoped that this classification will be used widely and without ad hoc modification, with resultant improvement in communication about patient status for both clinical and research purposes.

The working group was supported by funds provided by OMGE and Astra Hässle AB. Ola Jungardt is gratefully acknowledged for his statistical advice.

### Appendix

#### Endoscopic classification of reflux oesophagitis

<table>
<thead>
<tr>
<th>Institution</th>
<th>Investigator, Initials</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patient no</td>
<td>Country</td>
<td>Expert: Yes/No</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Trainee: Yes/No</td>
</tr>
</tbody>
</table>

#### Mucosal breaks

1. Is there a mucosal break?  
   If yes –
   - How many breaks < 5 mm diameter?  
     - Only one  
     - Many (2 or more)  
     - Do all have overlying white/yellow membrane
   - How many breaks > 5 mm diameter?  
     - Only one  
     - Many (2 or more)  
     - Do all have overlying white/yellow membrane

#### 2. How many folds are involved by mucosal breaks?

- Only one fold
- More than one fold
- Many folds (2 or more with mucosal breaks continuous between the tops of 2 or more folds, but not circumferential)
3. What proportion of the circumference is affected by mucosal breaks?

- 0–25% 3a. Yes/No
- 26–50% 3b. Yes/No
- 51–75% 3c. Yes/No
- 76–99% 3d. Yes/No
- 100% (circumferential) 3e. Yes/No

4. How many mucosal breaks are > 3 mm deep?

   - Only one 4a. Yes/No
   - Many (2+)

5. Is there an ulcer?

   - Only one 5a. Yes/No
   - Many (2+)

6. Is there an oesophageal stricture?

   - Only one 6a. Yes/No
   - If yes:
     - Is the stricture narrower than the endoscope? (i.e. does the endoscope fail to pass the stricture) 6b. Yes/No

7. Is there excessive reddening at the cardia? 7. Yes/No

8. Is columnar metaplasia (Barrett’s epithelium) present?

   - If yes:
     - does it form fingers
     - and/or islands? 8b. Yes/No

Complications

Stricture

Barrett’s oesophagus/metaplasia
Is it circumferential (> 3 cm extent)?

(8a and 8b are NOT mutually exclusive)

**Minor features suggestive of oesophagitis**

9. Is the distal oesophageal mucosa normal?

If it is not normal, is there:
- erythema in one segment at the squamocolumnar junction?
- blurring or indistinctness of the squamocolumnar junction?
- friability of the squamocolumnar junction?
- diffuse erythema in the distal oesophagus?
- patchy erythema in the distal oesophagus?
- increased vascularity in the distal oesophagus?
- oedema or accentuation of the mucosal folds?

**Overall summary of endoscopic features of oesophagitis**

10. Mark relevant features of oesophagitis on diagrams:

![Diagram of oesophagus with markings at different distances from 0 to 28 cm.](image-url)