The Endoscopic Assessment of Esophagitis: A Progress Report on Observer Agreement

DAVID ARMSTRONG,* JOHN R. BENNETT,† ANDRÉ L. BLUM,§ JOHN DENT,‖
F. TIMOTHY DE DOMBAL†,¶ JEAN–PAUL GALMICHE,# LARS LUNDELL,** MARIANA MARGULIES,¶
JOEL E. RICHTER,†† STUART J. SPECHLER,§§ GUIDO N. J. TYTGAT,‖‖ and LENE WALLIN¶¶

*Department of Medicine, Division of Gastroenterology, McMaster University Medical Center, Hamilton, Ontario, Canada; †Remenham House, West Ella, Hull, England; §Division de Gastro-entérologie, Centre Hospitalier, Universitaire Vaudois, Lausanne, Switzerland; ‖Gastroenterology Unit, Royal Adelaide Hospital, Adelaide, Australia; ¶Clinical Information Science Unit, Leeds, England; #Centre Hospitalier Regional et Universitaire de Nantes, Hôpital G & R Laënnee, Nantes, France; **Department of Surgery, University of Gothenburg, Sahlgren's Hospital, Gothenburg, Sweden; ††Division of Gastroenterology, Department of Medicine, Cleveland Clinic Foundation, Cleveland, Ohio; §§Department of Medicine, Beth Israel Hospital, Boston, Massachusetts; ‖‖Academisch Medisch Centrum, Darm-Enleverzeiken, Amsterdam, The Netherlands; and ¶¶Department of Surgical Gastroenterology, KAS, Glostrup, Denmark

Background & Aims: The study and management of reflux esophagitis require an endoscopic classification system founded on esophageal lesions that can be reproducibly identified. The aim of this study was to investigate interobserver agreement for the identification of endoscopic lesions typical of reflux esophagitis. **Methods:** Paired comparisons of observers' descriptions were obtained. Seventeen endoscopists assessed 100 still images, and 42 endoscopists, including 13 endoscopists in training, assessed 23 endoscopic video recordings. In a third, ancillary study, using a simpler evaluation sheet, 219 gastroenterologists recorded their assessments of 20 still images. **Results:** The agreement between endoscopists was similar for still images and video recordings. Agreement between experienced endoscopists was acceptable to good for recognition of minimal changes (erythema, friability, mucosal edema; $\kappa = 0.46$ to $\kappa = 0.8$), mucosal breaks (discretely, demarcated areas of slough or erythema; $\kappa = 0.84$), and complications (ulceration, $\kappa = 0.92$; stricturing, $\kappa = 0.80$; columnar metaplasia, $\kappa = 0.81$), although there was poor agreement when the circumferential extent and number of mucosal breaks were assessed. However, total circumferential extent of the mucosal break had a κ value of 0.59. Agreement between inexperienced endoscopists was poor for recognition of minimal changes but was good for recognition of complications (κ, 0.70–0.90). **Conclusions:** Endoscopists can identify mucosal breaks confined to a mucosal fold and lesions that extend throughout the esophageal circumference. Complications of reflux disease can be reproducibly recorded. Criteria for assessing the number of mucosal breaks and their radial extent must be defined more clearly, as must the features of minimal change esophagitis.

G astroesophageal reflux disease (GERD) is difficult to define because of its extreme heterogeneity. Peptic lesions of the squamous epithelium have a characteristic endoscopic appearance, but these lesions can occur transiently or even be absent in severely symptomatic patients.[1–5] Therefore, the endoscopic appearance has a high specificity, but lack of sensitivity, for GERD. Despite these limitations, it is quite clear that endoscopy remains the investigation of choice for making the diagnosis of reflux esophagitis and grading its severity.

The spectrum of endoscopic findings in reflux disease ranges from minimal, equivocal, mucosal changes through erosions to deep ulcerations, strictures, and columnar-lined esophagus. Data indicate that the clinical response to treatment and the subsequent prognosis are dependent on the severity of the mucosal lesions observed endoscopically.[6–12] Therefore, it is important to accurately classify the severity with which the peptic lesions affect the esophagus. With respect to esophagitis, the present situation is not that we lack endoscopic classification systems for severity but that a large number of such classifications exists.[13–21] A survey of existing work indicated that there are more than 30 such systems documented in the literature, none of which is universally accepted, each with individual proponents, many with overlapping criteria, and some mutually incompatible. This chaotic area is not likely to be helped by adapting an existing system or by developing another new system with the expectation that previous systems will be abandoned. Well-founded criticism has been directed against many existing classification systems, and difficulties are recognized in recommending any of them unconditionally. One important reason for this is that virtually all

---

†Deceased.

existing systems are based on visual observations that have not yet been assessed for reproducibility. To be widely used and accepted as a standard in clinical practice and also as a scientific tool, a system must fulfil a number of basic requirements. One of these is that the proposed criteria must be submitted to a validation program before they can be generally accepted for wide-spread clinical use. Our international working group has been engaged in problems related to the definition and evaluation of an endoscopic classification system for esophagitis, with the ultimate aim of its being universally accepted and found clinically useful. The initial aim was therefore to identify those endoscopic esophageal mucosal lesions that can be reproducibly described. We investigated the inter-observer variation in the identification and description of various lesions thought to characterize uncomplicated esophagitis, as well as complications of the disease, by an analysis of still images and video images.

Materials and Methods

Three interlinked, but separate, studies of observer agreement in the visual appearances of GERD at endoscopy were performed.

Still Image Analyses

Seventeen observers from three countries assessed the reproducibility in recordings of specific endoscopic appearances from a series of color slides. Each participant was given a set of 100 slides from those in the possession of one of the authors (G.N.J.T.), showing visual endoscopic features of GERD ranging from minimal change to severe morphological damage. Data from each patient were recorded on a predesigned work sheet (Figure 1).

Video Image Analyses

The video image study involved 42 observers from seven countries. The objective was to assess whether the differences shown in the first study were influenced by the artificiality of still images used in the slide presentations. Each participant was given video tapes, recorded using Olympus video endoscopes and processed on Super VHS or NTS Video Recording Systems, showing endoscopic appearances from 23 different patients with GERD of varying severity. The recording from each patient lasted approximately 30 seconds, and the same work sheet was used as in the previous study (Figure 1). The experiences of the observers varied from endoscopists under training (who had performed less than 500 upper gastrointestinal endoscopies) to experts (who had performed more than 3000 upper gastrointestinal endoscopies). Of the 42 participants, 29 were experts.

Mucosal Break Analysis

A proposal for a new endoscopic classification, with emphasis on mucosal breaks, was presented to a symposium at the World Congress of Gastroenterology in Los Angeles in 1994. Mucosal breaks were graded as A, B, C, and D according to the extent (Table 1). After an initial presentation of the rationale for and content of the proposed endoscopic classification, 219 delegates recorded their impression of 20 still images.

Analysis of Data

A total of 103,000 data points were recorded from the still image and video image studies. In terms of comparison between observers, a total of 1994 pairs of observers were potentially available from the two studies alone. However, because comparisons were made within groups (e.g., each expert was compared with each other expert but not with each endoscopist under training), the total number of two-way comparisons made was 872 between pairs of observers.

The analyses of the large volume generated by these studies were performed in Leeds, England, using the facilities of the Clinical Information Science Unit. The 103,000 data points were analyzed using a 386 desk-top computer via the dBASE IV program. Statistical analyses were performed based on the κ statistics of Cohen.[22] For those unfamiliar with the possible range of this statistic, Figure 2 shows the interval within which this κ statistic usually falls, along with the customary interpretation of the κ findings. For each of the appearances in Figure 1, the κ values between each pair of observers were calculated and the results were expressed in Figures 3–6 as mean kappa values and an interquartile range.

Data from the (mucosal break) study at the Los Angeles World Congress were analyzed using different (simpler) methods. Similar (κ) analyses would have resulted in about 48,000 two-way comparisons between widely disparate observers. The results are therefore presented as the proportion of still images in which 90% of those delegates observing the image agreed that (according to the criteria in Table 1) the appearance was or was not present. For details of statistical analyses and further statistical comment, see the Appendix.

Results

This section refers to the still image and video image studies. A subsequent added note concerns the study conducted (under less controlled conditions) as the Los Angeles World Congress.

Severe Complications

Severe complications of reflux disease (ulcer, stricture, or columnar-lined mucosa)[23,24] were noted by the majority of observers in approximately one quarter of the observations made: ulcer craters were claimed to be present in 10% of the responses concerning video images and 15% of the responses concerning still images. These findings of the studies involving video images were broadly similar.

These changes were reliably reported (Figure 3). Among experienced endoscopists reviewing still images, κ values were high for ulcer (0.92), stricture (0.8), or columnar-lined

Example No _ _ _   Investigator No _ _ _   Centre No _ _ _ _ _ _ _Date _ _ / _ _ / _ _

Experienced endoscopist ☐   Endoscopist under training ☐

		MINOR CHANGES	
	A1	Localised area(s) of erythema in one or more segments at mucosal junction	
	A2	Indistinctness or blurring of all or part of mucosal junction	
	A3	Friability at the mucosal junction	
	A4	Diffuse erythema in distal oesophagus	
	A5	Patchy erythema in distal oesophagus	
	A6	Increased vascularity in distal oesophagus	
	A7	Oedema / accentuation of mucosal folds	
		MUCOSAL BREAKS (discrete areas of erythema ± exudate)	
	B1	≤ 5 mm along longest axis - **number** with no overlying yellow exudate	
	B2	- **number** with yellow exudate over all or part of lesion	
	B3	> 5 mm along longest axis - **number** with no overlying yellow exudate	
	B4	- **number** with yellow exudate over all or part of lesion	
	C1	**Number** located only on the apex of a fold	
	C2	**Number** located only between folds	
	C3	**Number** extending from one fold to the next	
	C4	**Number** whose location is uncertain	
	D1	1 fold only involved	
	D2	≥ 2 folds involved but not circumferential	
	E1	≤ 25% of circumference involved	
	E2	> 25%, but < 100% of circumference involved	
	F1	Circumferential	
	G1	ULCERS - Mucosal break ≥ 2 mm deep + overlying exudate (**Number**)	
	G2	- Contiguous with mucosal breaks extending around rest of oesophagus	
	H1	STRICTURE Absent	
	H2	Present > diameter of endoscope	
	H3	≤ diameter of endoscope	
	I1	METAPLASIA Absent	
	I2	Present (continuous with mucosal junction) "fingers" ± "islands"	
	I3	circumferential (≥ 3 cm)	

Figure 1. Sample of work sheet used in still image and video image analyses.

esophagus, with comparable κ values (0.81, 0.83, and 0.84, respectively) for experienced endoscopists' observations from the series of video images. Observations made by endoscopists under training were slightly less homogeneous but still well within limits of acceptability, given the conventional interpretation of κ statistics (ulcer, 0.9; stricture, 0.7; and columnar-lined esophagus, 0.7).

Mucosal Breaks

The presence or absence of a mucosal break on a particular still or video image (Figure 4) was reliably recorded by all grades of endoscopists (mean κ values, 0.81–0.84). In terms of the extent of mucosal breaks, the data from these studies indicated less agreement.

Table 1. The Los Angeles Classification of Esophagitis

Grade A	One or more mucosal breaks confined to the mucosal folds, each no longer than 5 mm
Grade B	At least one mucosal break more than 5 mm long confined to the mucosal folds but not continuous between the tops of two mucosal folds
Grade C	At least one mucosal break continuous between the tops of two or more mucosal folds but not circumferential
Grade D	Circumferential mucosal break

Figure 3. κ values for comparisons made among experienced endoscopists in recognizing severe esophagitis changes (ulcer, stricture, and columnar-lined mucosa). In Figures 3–6, horizontal bars represent mean of all two-way interobserver comparisons and columns represent interquantile range. ■, Video; □, slides.

Comparisons between experienced endoscopists evaluating video images resulted in κ values of about 0.4 for most aspects concerning the length of break and the presence or absence of yellow exudate (see B1 to B4 in Figure 1). Similar values emerged from observations of still images. One value was unusual: feature B2 (small mucosal breaks with yellow exudate over all or part of the lesion) was recorded with excellent agreement between observers. However, it should be noted that this lesion was infrequently seen (only noted in 3%).

In terms of circumference, two categories were reliably identified. These were categories D1 (total involvement was equal to or less than onefold) and F1 (circumferential involvement), with κ values of 0.84 and 0.59, respectively. Distinction between intermediate categories of break proved difficult (κ, <0.4).

Finally, attempts to calculate the number of mucosal breaks (C1 to C4 in Figure 1) resulted in completely unreliable figures, ranging (in the same individual patient) from 1 to 15 on several occasions. These data were so obviously variable that they were not subjected to formal analysis.

Minimal Changes

A different pattern emerged from analyses of minimal changes either in the distal esophagus or at the squamocolumnar junction (section A in Figure 1) (Figures 5 and 6). Distinct differences were shown between
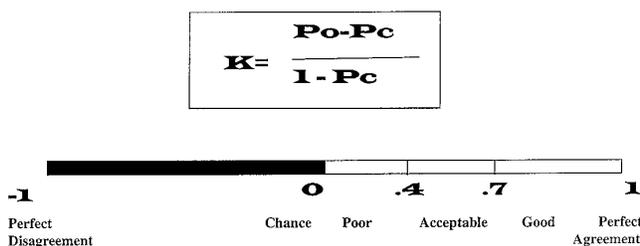
the reproducibility of observations made by experienced endoscopists and endoscopists under training. Virtually all minimal distal and mucosal junction changes (Figure 1) were reproducibly recorded by experienced endoscopists, with κ values around 0.8 for increased vascularity, local erythema, and friability. By contrast, corresponding values for endoscopists under training were significantly lower, with the highest κ value of 0.39 for increase in vascularity and with values between 0.19 and 0.22 for blurring of the mucosal junction, friability, and edema.

Los Angeles Study

A study was performed at the symposium of the World Congress of Gastroenterology in Los Angeles examining the agreement between 219 observers grading 20 exposed slide images. Small mucosal breaks (<5 mm in extent) graded as A were reliably recorded, with a κ value of 0.65. There was also a good agreement about the presence or absence of circumferential breaks (grade D) attaining a κ of 0.55. In contrast, there was less agreement concerning intermediate grading of breaks (grade B and C, 0.30 and 0.10, respectively). Comparing these findings and the κ values from the more formal studies (Figures 3–6) showed astonishingly similar outcomes.

Figure 2. Explanation of κ statistics of Cohen[22] showing possible rates of values and conventional interpretation of results. Po is the observed proportion of agreement and Pc is the expected (change) agreement in the relevant contingency table.
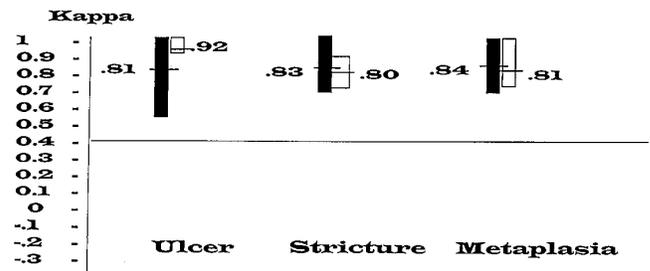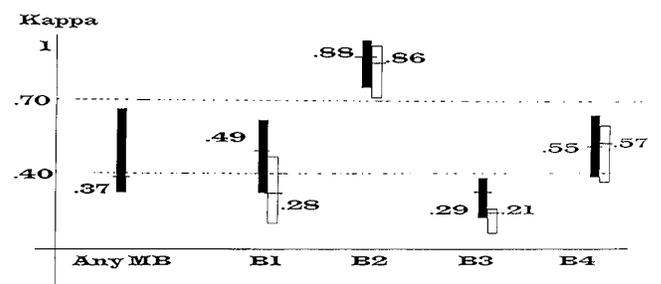

Figure 4. Agreement concerning mucosal breaks (MB) assessed on video images by experienced endoscopists (■) and endoscopists under training (□).
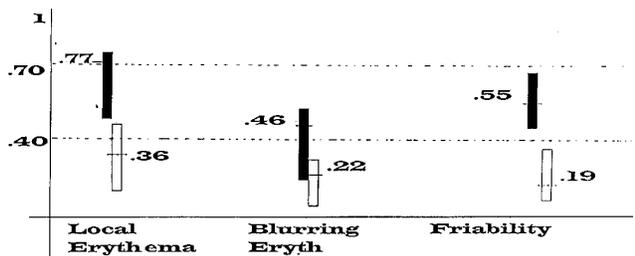
Figure 5. Analysis concerning minimal changes at mucosal junction on video images evaluated by experienced endoscopists (■) and endoscopists under training (□).

Discussion

A useful endoscopic classification of reflux disease must fulfil a number of criteria. It should identify lesions specific for GERD with high accuracy and minimal interobserver variability. The identification of lesions should not be dependent on subtle nuances of fiberoptic endoscopic technology, and the classification should minimize the problems of misinterpreting the extent of the lesion(s) either longitudinally or horizontally. Furthermore, it should describe each lesion exactly, irrespective of the coexistence of other lesions. An important feature is to allow a practical collection of data and description of all concomitantly existing lesions, and each grade of esophagitis should be distinct enough to avoid overlap. It is also important that the endoscopic classification can be easily memorized and provide a standardized, generally accepted, and comprehensive description of the disease specific features while differentiating clearly among minimal changes, mucosal breaks, and complications. The system should also distinguish reversible from irreversible lesions.

The term mucosal break was introduced with the hope of avoiding confusion in the use of the terms erosion and ulceration. These latter terms are primarily histological rather than visual and imply a greater degree of certainty about the judged depths of a mucosal lesion than may be possible through an endoscope. Furthermore, the terms erosion and ulceration seem to be used in different ways by different observers.[6,13–21] Yet, for any assessment of extent, a clear and pragmatic working definition of the endoscopic appearance of the mucosal break is mandatory. In the working group, some agreement was achieved on the definition of a mucosal break: an area of slough or an area of erythema with a discrete lined demarcation from the adjacent or normal looking mucosa. We decided that the peaks of mucosal folds should be reference points, and we considered that these could be identified during partial air inflation of the esophagus. The scoring

of extent was designed to be as simple as possible but with adequate discrimination between the degrees of severity of esophagitis for clinical and research purposes, regardless of esophageal length. It is obviously important to determine the extent of mucosal lesions independent of other measures of severity because the extent of esophagitis may be the single most important measure to define severity and the need for subsequent therapy.[6,7,12,14,25–28] It could be argued that the smallest mucosal break that can be identified is a relatively minor abnormality, not carrying any significant risk for the patient. A similar lesion extending to more than 5 mm might be associated with a greater amount of acid exposure of the squamous epithelium. Similarly, the radial extent of mucosal breaks might also be related to severity of reflux.[29–31]

This study, using both still and video images to assess the interobserver variability with regard to disease-specific lesions affecting the esophageal mucosa, is unique primarily because of the magnitude of the study and the huge number of comparisons made. The agreement between observers was acceptable when determining the presence or absence of a mucosal break confined to a mucosal fold. In addition, extensive lesions, involving the entire circumference of the esophagus, could be determined with high accuracy. Furthermore, complications of the disease, such as the presence of strictures, ulcers, and Barrett's columnar-lined mucosa, were assessed with extremely good agreement between observers, essentially irrespective of whether they were experienced endoscopists or under training. Other studies have reported markedly lower κ values for the overall endoscopic grading of esophagitis.[15,32,33] Differences in design and size of the studies may explain much of the apparent differences in outcome. In general terms, by using video recordings, it is possible that an underestimation of the true κ values will occur. Video recordings are different from standard endoscopic procedures, especially because the observers do not have the benefit of moving the endoscope through the organ.[15] From our results, it can be concluded that single mucosal breaks confined to the
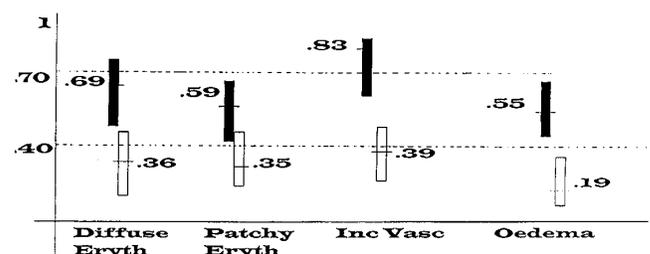

Figure 6. Same analysis as in Figure 3 concerning minimal changes in distal esophagus for comparisons between experienced endoscopists (■) and endoscopists under training (□).
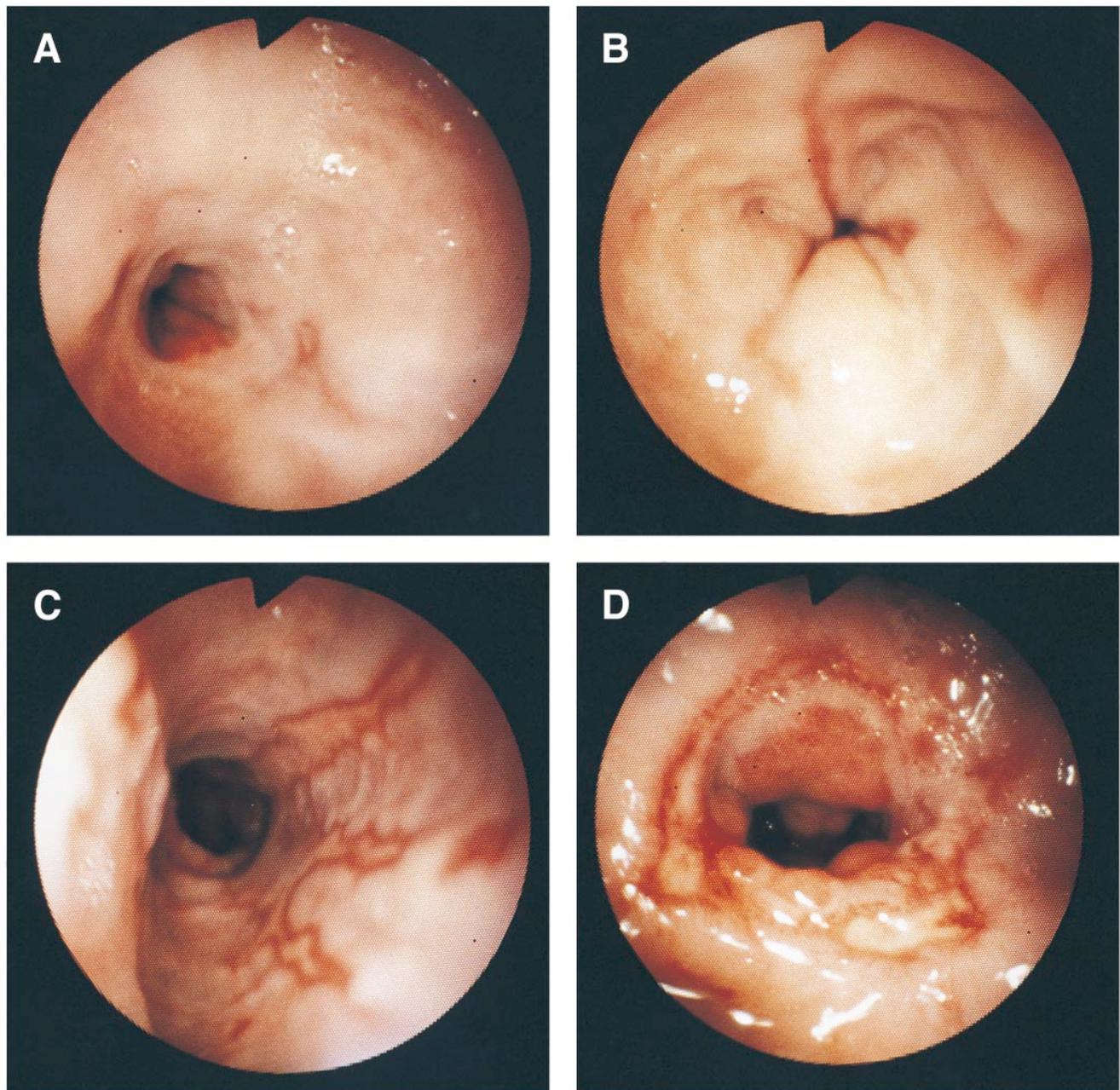
Figure 7. (*A*) Mucosal breaks confined to the mucosal fold, each no longer than 5 mm. (*B*) At least one mucosal break longer than 5 mm confined to the mucosal fold but not continuous between two folds. (*C*) Mucosa breaks that are continuous between the tops of mucosal folds but not circumferential. (*D*) Circumferential mucosal break with one portion being of significant depth.

mucosal fold, as well as mucosal breaks involving most of the esophageal circumference, can be accurately and reproducibly assessed. The question remains how to interpret the low κ values for mucosal breaks involving two or more folds and the valleys in between. It is pertinent to emphasize that the design of the work sheet used in the present evaluation program was aimed at gaining basic information on different endoscopic features rather than specifically focusing on the accuracy by which the radial extent of mucosal breaks could be assessed. At this stage,

it is essential to design further studies specifically focusing on the radial extent of esophagitis using refined video images and a work sheet individually designed to address this issue.

The inclusion of "minimal lesions" in the endoscopic diagnosis of GERD has been found to improve the sensitivity of endoscopy compared with that of other diagnostic tests.[13,34] The improvement in sensitivity, however, can only be achieved at the expense of a markedly decreased specificity. In fact, studies have previously found

no correlation between the subtle endoscopic changes of "grade 1 esophagitis" and histological changes implying GERD.[5,13] Patients with mild or minimal lesions also have a weaker association between symptoms and documented reflux episodes (as assessed by esophageal pH-metry) than those with more clear-cut evidence of esophagitis.[1,13,34,35]

It might be concluded from reviewing the literature that these "minor" endoscopic stigmata represent disputable endoscopic lesions, resulting in great discrepancies in the role of minimal changes in the diagnosis of the disease. In the present study, we obtained unexpectedly high κ values among experienced endoscopists with respect to minimal changes. This observation is also unexpected in view of previous data reporting a low level of agreement when similar patients were concomitantly assessed by different investigators during "live endoscopy."[15] However, previous studies did not measure κ values for each individual endoscopic finding comprising the entity "minimal changes," whereas the present study has done so. The drawbacks and pitfalls of κ statistics in, for instance, the absence of an item have also to be recognized.[36] We nevertheless consider it important to obtain additional information on interobserver variability when this minimal change issue is specifically addressed in the next phase of this evaluation process. Perhaps modern endoscopic technologies will allow more refined video images to be presented, offering the potential for higher accuracy and resolution in the assessment.

A proposal for a new endoscopic classification system with emphasis on mucosal breaks, graded A to D according to the extent, was presented to a symposium at the World Congress of Gastroenterology in Los Angeles in 1994 (Table 1 and Figure 7). At this symposium, 219 delegates recorded their impression of 20 still images. The comparison between their results and the κ values from the more formal studies showed good agreement despite the different manner of conducting them. Small breaks (grade A) were reliably recorded in both studies. Equally, there was a good agreement in both studies about the presence or absence of circumferential breaks. In constrast, there was less agreement concerning intermediate grades. The results of the set of studies thus strongly support the potential usefulness of grades A and D with some additional intermediate category. The precise details and definition of this intermediate category need to be further developed.

It must be emphasized that this endoscopic classification system, which still represents a working model, is not yet ready for clinical application. Adaptation of an incompletely developed scheme would result in its abandonment, which has happened to many existing systems.

Forthcoming studies, already in progress, have to determine the interobserver agreement on the radial extent of mucosal breaks. Studies are also under way to determine the usefulness of the proposed endoscopic classification in clinical practice. In this context, the proposed endoscopic classification will be compared with the commonly used Savary–Miller classification.[19,20] In addition, the predictive value of the proposed grading in the subsequent short- and long-term outcome on medical therapy in GERD has to be determined.

Appendix

Statistical Analysis

Despite the recognition for at least 40 years that observer variation is important in clinical medicine, there is still unfortunately no absolute criterion by which observer variation in clinical medicine can be measured. The most widely used coefficient of agreement in clinical studies is the κ statistic of Cohen.[22]

For the purposes of this study, for each individual characteristic in Figure 1, we analyzed data from a large number of observers as a multiple of comparisons between each pair of observers. That is, if 20 observers recorded the presence or absence of a particular feature, κ statistics were calculated for a total of (n × n − 1) (20 × 19) (380) comparisons. In this way, each observer is compared with each other observer.

We used the κ statistic in its original version. P_o is the observed proportion of agreement and P_c is the expected agreement by chance in the relevant contingency table.

The range of possible values for κ and the conventional interpretation of the κ statistic are shown in Figure 2. In Figures 3–6, the data from the present study are described in a simple fashion. For each attribute, the pooled observer agreement in recording that attribute is described by way of a mean κ value (illustrated by a bar on the Figure) for all two-way interobserver comparisons together with an upper and lower quartile band (illustrated by a column). Where appropriate, observer subgroups are specified in the text and/or legends to Figures.

Finally, it needs to be stressed that the study represents an interobserver study, not an intraobserver study.

References

1. Galmiche JP, Bruley des Varannes S. Symptoms and diseases severity in gastro-oesophageal reflux disease. Scand J Gastroenterol 1994;29:201:62–68.
2. Pace F, Santalucia F, Bianchi Porro G. Natural history of gastro-oesophageal reflux disease without oesophagitis. Gut 1992;32:858–848.
3. Schindlbeck NE, Klauser AG, Berghammer G, Londong W, Müller-Lissner SA. Three year follow up of patients with gastro-oesophageal reflux disease. Gut 1992;33:1016–1019.
4. Spechler SJ. Epidemiology and natural history of gastro-oesophageal reflux disease. Digestion 1992;51:(Suppl 1):24–29.
5. Armstrong D, Emde C, Inauen W, Blum AL. Diagnostic assessment of gastroesophageal reflux disease: what is possible vs what is practical? Hepatogastroenterology 1992;39:3–13.

6. Hetzel DJ, Dent J, Reed WD, et al. Healing and relapse of severe peptic esophagitis after treatment with omeprazole. Gastroenterology 1988;95:903–912.

7. Sandmark S, Carlsson R, Fausa O, Lundell L. Omeprazole or ranitidine in the treatment of reflux esophagitis. Results of double-blind, randomized, Scandinavian multicentre trial. Scand J Gastroenterol 1988;23:625–632.

8. Tytgat GN, Nio CY, Schotborgh RH. Reflux esophagitis. Scand J Gastroenterol 1990;25:(Suppl 175):1–12.

9. Koelz HR, Birchler R, Bretholz A, et al. Healing and relapse of reflux oesophagitis during treatment with ranitidine. Gastroenterology 1986;91:1198–1205.

10. Tytgat GNJ, Hansen OJA, Carling L, Degroot GH, Geldof H, Glise H, Elfskind P, Elsborg L, Karvonen AL, Ohlin B, Solhaug OH, Vermeersch B. Effect of cisapride on relapse of reflux oesophagitis, healed with an antisecretory drug. Scand J Gastroenterol 1992;27:175–183.

11. Lundell L, Backman L, Ekström T, et al. Prevention of relapse of reflux esophagitis after endoscopic healing: the efficacy and safety of omeprazole compared with ranitidine. Scand J Gastroenterol 1991;26:248–256.

12. Dent J, Yeomans ND, Mackinnon M, Reed W, Narielvala FM, Hetzel DJ, Solcia E, Shearman DJS. Omeprazole versus ranitidine for prevention of relapse in reflux oesophagitis. A controlled double-blind trial of the efficacy and safety. Gut 1994;35:590–598.

13. Armstrong D, Monnier P, Nicolet M, Blum AL, Savary M. Endoscopic assessment of oesophagitis. Gullet 1991;1:63–67.

14. American Society of Gastrointestinal Endoscopy. The role of endoscopy in the management of esophagitis—guidelines for clinical application. Gastrointest Endosc 1988;34:(Suppl 9s).

15. Bytzer P, Havelund T, Möller Hansen J. Interobserver variation in the endoscopic diagnosis of reflux esophagitis. Scand J Gastroenterol 1993;28:119–125.

16. Day DW. Infective, physical and chemical oesophagitis. In: Whitehead R, ed. Gastrointestinal and oesophageal pathology. Edinburgh: Churchill Livingstone 1989:377–383.

17. Giuli R, Savary M, Skinner DB. Esophagitis and Barrett's esophagus. In: Collaborative studies on esophagitis and esophageal dysplasia. Paris: Organisation Internationale d'Etudes Statistiques pour les Maladies de l'Oesophage, 1989.

18. Makuuchi H, Mitomi T. A new endoscopic criterion and classification of reflux esophagitis. Abstracts from the 42nd Congress of the Japan Gastroenterological Endoscopy Society (abstr). Dig Endosc 1992;4:289.

19. Monnier P, Savary M. Contribution of endoscopy to gastro-oesophageal reflux disease (abstr). Scand J Gastroenterol 1984;106:26.

20. Ollyo JB, Gontollier C, Brossard E, Lang F. La nouvelle classification de Savary des oesophagites de reflux. Acta Endosc 1992;22:307.

21. Tytgat GN, Molenaar CBH. Grading endoscopic severity in gastroesophageal reflux disease. Rom J Gastroenterol 1994;3:91–93.

22. Cohen J. A coefficient of agreement for nominal scales. Educ Psychol Measurement 1960;20:37–46.

23. Spechler SJ. Complications of gastroesophageal reflux disease. In: Castell DO, ed. The esophagus. Boston: Little Brown, 1992:543–556.

24. Dent J, Bremner CG, Collens MJ, et al. Barrett's oesophagus.

Working party report to the World Congress of Gastroenterology, Sydney 1990. J Gastroenterol Hepatol 1991;6:1–22.

25. Bennett JR, Chapman RWG, Harvey RF, Heading RC, Morris AI, Stevens RM. Management guidelines for gastro-oesophageal reflux disease. The European Digestive Disease Week, Amsterdam, 1991:46–54.

26. Lundell L. Long-term treatment of gastro-oesophageal reflux disease with omeprazole. Scand J Gastroenterol 1994;29:(Suppl 201):74–78.

27. Blum AL, Adami B, Bouzo MH, Brandstätter G, Fumagalli I, Galmiche JP, Hebbelin H, Heutschel E, Hüttemann W, Schütz E, Verlinden M, Italian Eurocis Trialists. Effect of cisapride on relapse of esophagitis. A multinational placebo-controlled trial in patients healed with an antisecretory drug. Dig Dis Sci 1993;38:551–560.

28. Kuster E, Ros E, Toledo-Pimentel V, Pujol A, Bordas JM, Grande J, Pera C. Predictive factors of the long-term outcome in gastro-oesophageal reflux disease: six year follow up of 107 patients. Gut 1994;35:8–14.

29. Johnsson F, Joelsson B, Isberg P-E. Ambulatory 24 hour intra-esophageal pH-monitoring in the diagnosis of gastroesophageal reflux disease. Gut 1987;28:1145–1150.

30. Singh S, Taylor RH, Colin-Jones DG. Simultaneous two level oesophageal pH monitoring in healthy controls and patients with oesophagitis: comparison between two positions. Gut 1994;35:304–308.

31. Baldi F, Ferrarini F, Langanesi A, Ragazzini M, Barbava L. Acid gastro-oesophageal reflux and symptom occurrence. Analyses of some factors influencing their association. Dig Dis Sci 1989;34:1890–1893.

32. Gustavsson S, Bergström R, Erwall C, Krog M, Lindholm C-E, Nyrén O. Reflux esophagitis: assessment of therapy effects and observer variation by video documentation of endoscopy findings. Scand J Gastroenterol 1987;22:585–591.

33. Kling PA, Edin K, Domellöf L. Observer variability in upper gastrointestinal fiberendoscopy. Scand J Gastroenterol 1985;20:462–465.

34. Schüle A, Brändli H, Belloni S, Koelz HR, Pirozynski WJ, Blum AL. Endosckopische Diagnose der Öesophagitis. Wo liegt die Grenze zum normalen? Deutsche Medicinishe Wochenschrift 1977;102:606–609.

35. Havelund T, Laursen LS, Lauritsen K. Efficacy of omeprazole in lower grades of gastro-oesophageal reflux disease. Scand J Gastroenterol 1994;29:201:69–73.

36. Lind T, Havelund T, Carlsson R, Glise H, Hernqvist H, Junghard O, Lauritsen K, Lundell L, Pedersen SA. The effect of omeprazole (OME) 20 mg daily on heartburn in patients with endoscopy negative reflux disease (ENRD) (abstr). Scand J Gastroenterol 1995;108:A151.

37. Brennan P, Silman A. Statistical methods for assessing observer variability in clinical measures. BMJ 1992;304:1491–1494.