

Interobserver agreement in endoscopic evaluation of reflux esophagitis using a modified Los Angeles classification incorporating grades N and M: A validation study in a cohort of Japanese endoscopists

H. Miwa,¹ T. Yokoyama,² K. Hori,¹ T. Sakagami,¹ T. Oshima,¹ T. Tomita,¹ Y. Fujiwara,³ H. Saita,⁴ T. Itou,⁵ H. Ogawa,⁶ Y. Nakamura,⁷ K. Kishi,⁸ Y. Murayama,⁹ E. Hayashi,¹⁰ K. Kobayashi,¹⁰ N. Tano,¹¹ K. Matsushita,¹¹ H. Kawamoto,¹² Y. Sawada,¹³ A. Ohkawa,¹⁴ E. Arai,¹⁵ K. Nagao,¹⁶ N. Hamamoto,¹⁷ Y. Sugiyasu,¹⁸ K. Sugimoto,¹⁹ H. Hara,²⁰ M. Tanimura,²¹ Y. Honda,²² K. Isozaki,²³ S. Noda,²⁴ S. Kubota,²⁵ S. Himeno²⁶

¹Department of Internal Medicine, Hyogo College of Medicine, Nishinomiya, ²Department of Technology Assessment and Biostatistics, National Institute of Public Health, Waco, ³Second Department of Surgery, Hyogo College of Medicine, Nishinomiya, ⁴Hyogo Prefectural Amagasaki Hospital, Nishinomiya, ⁵Kansai Rosai Hospital, Amagasaki, ⁶Nishinomiya Municipal Central Hospital, Nishinomiya, ⁷Kyowa Marina Hospital, Nishinomiya, ⁸Meiwa General Hospital, Nishinomiya, ⁹Itami City Hospital, Itami, ¹⁰Kinki Central Hospital of Mutual Aid Association of Public School Teachers, Amagasaki, ¹¹Kyoritsu Hospital, Kawanishi, ¹²Kawanishi Municipal Hospital, Kawanishi, ¹³Takarazuka Municipal Hospital, Takarazuka, ¹⁴Higashi Takarazuka Satoh Hospital, Takarazuka, ¹⁵Nishinomiya Kyoritsu Neurosurgical Hospital, Nishinomiya, ¹⁶Nagao Clinic, Amagasaki, ¹⁷Hamamoto Clinic, Amagasaki, ¹⁸Sugiyasu Clinic, Amagasaki, ¹⁹Sugimoto Kenji Clinic, Suita, ²⁰Hara Surgery Clinic, Nishinomiya, ²¹Tanimura Gastro-Enterological Clinic, Takarazuka, ²²Doctor Honda's office, Itami, ²³Isozaki Clinic, Kawanishi, ²⁴Noda Clinic, Nishinomiya, ²⁵Kano Hospital, Osaka, ²⁶Ashiya Municipal Hospital, Ashiya, Japan

SUMMARY. The Los Angeles classification system is the most widely employed criteria associated with the greatest interobserver agreement among endoscopists. In Japan, the Los Angeles classification system has been modified (modified LA system) to include minimal changes as a distinct grade of reflux esophagitis, rather than as auxiliary findings. This adds a further grading M defined as minimal changes to the mucosa, such as erythema and/or whitish turbidity. The modified LA system has come to be used widely in Japan. However, there have been few reports to date that have evaluated the interobserver agreement in diagnosis when using the modified LA classification system incorporating these minimal changes as an additional grade. A total of 100 endoscopists from university hospitals and community hospitals, as well as private practices in the Osaka-Kobe area participated in the study. A total of 30 video clips of 30–40 seconds duration, mostly showing the esophagocardiac junction, were created and shown to 100 endoscopists using a video projector. The participating endoscopists completed a questionnaire regarding their clinical experience and rated the reflux esophagitis as shown in the video clips using the modified LA classification system. Agreement was assessed employing kappa (κ) statistics for multiple raters. The κ -value for all 91 endoscopists was 0.094, with a standard error of 0.002, indicating poor interobserver agreement. The endoscopists showed the best agreement on diagnosing grade A esophagitis (0.167), and the poorest agreement when diagnosing grade M esophagitis (0.033). The κ -values for the diagnoses of grades N, M, and A esophagitis on identical video pairs were 0.275–0.315, with a standard error of 0.083–0.091, indicating fair intraobserver reproducibility among the endoscopists. The study results consistently indicate poor agreement regarding diagnoses as well as fair reproducibility of these diagnoses by endoscopists using the modified LA classification system, regardless of age, type of practice, past endoscopic experience, or current workload. However, grade M reflux esophagitis may not necessarily be irrelevant, as it may suggest an early form of reflux disease or an entirely new form of reflux esophagitis. Further research is required to elucidate the pathophysiological basis of minimal change esophagitis.

KEY WORDS: endoscopy, interobserver agreement, kappa value, LA classification, minimal change, reflux esophagitis.

INTRODUCTION

A number of classifications of reflux esophagitis have been proposed to date,¹ with the Los Angeles classification system (LA system)² being the most widely employed criteria and generating the greatest interobserver agreement among endoscopists.²⁻⁵ Of note, the LA system has ruled out minimal changes such as erythema, increased vascularity, friability, and edema as not consistently detectable, to increase interobserver agreement by limiting the grading to erosions or erosive esophagitis.^{2,3} However, of these minimal changes, erythema or increased vascularity is reported to lend itself far more consistently than edema or friability to endoscopic detection. The kappa (κ) values among experienced endoscopists are reported to be between 0.59 and 0.83, indicating good agreement, while the κ -values among inexperienced endoscopists are reported to have a range of 0.36–0.39, indicating fair agreement.²

In Japan, the Los Angeles classification system has been modified (modified LA system) to include minimal changes as constituting a distinct grade of reflux esophagitis, rather than as auxiliary findings, thus adding further grades of N (defined as normal mucosa) and M (minimal changes to the mucosa such as erythema and/or whitish turbidity).^{1,6} The modified LA system has become widely used among Japanese endoscopists, employing this system for the evaluation of reflux esophagitis. Furthermore, the majority of endoscopically detected minimal changes are reported to be whitish ones.¹ Whitish mucosa was considered a characteristic of acid-induced mucosal injury. Micro injuries of esophageal epithelium by excessive acid reflux may result in epithelial thickening, which leads to the whitish turbidity of mucosa. Histological features of whitish mucosa have been reported as the lengthening of papillae and hyperplasia of the basal zone^{7,8} or acanthosis with or without epithelial keratinization,⁹ which, in turn, are interpreted as representing hyperplastic changes present in the mucosa as a consequence of acid reflux, and therefore a histological finding suggestive of the presence of mild or nonerosive reflux esophagitis (NERD).

As recognition of these changes draws on the ability of the endoscopist to differentiate findings by color, it is highly likely that there exists wide variability among endoscopists connecting the diagnosis of these lesions. Accordingly, whether these findings are recognized correctly and consistently will have an impact on clinical practice. However, few studies have examined interobserver variations in the diagnosis of this minimal change esophagitis.¹⁰ Therefore, the clinical significance of this modified LA classification system remains invalidated.

MATERIALS AND METHODS

Participating endoscopists

A total of 100 endoscopists from university and community hospitals, as well as private practices in the Osaka-Kobe area participated in the study. Prior to the study, these participants were asked to undergo a questionnaire survey to obtain information on their age, sex, type of practice, specialty, number of years in endoscopic practice, number of endoscopic procedures performed, institutions where they were trained as endoscopists, their weekly endoscopic workload, interest in reflux esophagitis, whether or not they often make a diagnosis of grade M reflux esophagitis (color change-based grading) or hiatal hernia, and what they regard as the hallmark of Barrett's esophagus.

Preparation of video clips to be used for the diagnosis of reflux esophagitis

Videotapes were randomly selected from those generated during the endoscopic examinations performed by two of the authors (HM and TS) in patients with normal to mild reflux esophagitis at a certain clinic or hospital on a particular day, and a total of 30 video clips of 30–40 seconds duration that mostly featured the esophagocardiac junction were then created. Three of these video clips were randomly selected and duplicated so that the endoscopists would be shown three pairs of identical clips during their evaluation; the endoscopists were not informed of this.

Evaluation of video clips

The video clips were shown on a single large screen to 100 endoscopists using a video projector, which was large enough for all of the participants to see the images clearly. The participating endoscopists were instructed beforehand to grade the reflux esophagitis for each of the cases using the modified Los Angeles classification (incorporating grades N and M) that is commonly used in Japan (Table 1-a).^{1,6} All participants were assembled in a single conference room to watch the video clips just once on a single large screen, and they marked their diagnosis on separate sheets independently. Hereby, they watched them under the same conditions and their diagnoses would not have been affected by the judgment of others. They were also asked to evaluate the presence or absence of Barrett's epithelium or hiatal hernia during their video-based diagnoses, as well as to rate the level of difficulty in making a diagnosis using a visual analog scale for each of the video clips presented.

Statistical analysis

The frequency distributions of categorical data are summarized as percentage values. Interobserver

Table 1-a A modification of the Los Angeles classification system used in Japan

Grade	Description
N	Normal mucosa
M	Minimal changes to the mucosa, such as erythema and/or whitish turbidity
A	Non-confluent mucosal breaks < 5 mm in length
B	Non-confluent mucosal breaks > 5 mm in length
C	Confluent mucosal breaks < 75% circumferential
D	Confluent mucosal breaks > 75% circumferential

Table 1-b General interpretation of kappa value

0	Poor
0–0.2	Slight
0.2–0.4	Fair
0.4–0.6	Moderate
0.6–0.8	Substantial
0.8–1.0	Almost perfect

agreement in the diagnoses of esophagitis (3 grades: N, M, and A), hiatal hernia, and Barrett's epithelium was expressed in terms of the kappa value for multiple raters,¹¹ which gives an estimate of agreement in excess of that predicted by chance alone (Table 1-b). Kendall's coefficient of concordance was also calculated to assess interobserver agreement considering the order of N, M, and A esophagitis. The intraobserver reproducibility of the endoscopic diagnoses was evaluated in terms of the κ -value for two readings in each of the three duplicated video clips. The differences of two κ -values were statistically tested by the Z-test based on standard errors of κ -values. Four of the 100 participants who did not evaluate three or more video clips were excluded from all analyses. The κ -values were calculated using data of 96 participants who completed readings for all 33 video clips. All statistical analyses were done with the use of the SAS statistical package (version 9.1; SAS Institute, Cary, NC, US) and MAGREE macro for SAS.

RESULTS

Participant background

The participants ($n = 96$) ranged in age from their 20s to 70s, with the majority being in their 40s, followed by those in their 30s, and constituted a highly male-dominated group, with 92.6% of the participants being men. The majority of participants (66.3%) were community hospital physicians, followed by general practitioners (20%), and university hospital physicians (13.7%). In terms of specialty, 80% of the endoscopists were internists, with 15.8% being surgeons. Their experience as endoscopists varied from less than 5 years to more than 20 years,

with 46.3% of the participants having 11–20 years, and 24.2% having 5–10 years of experience. The number of endoscopic procedures performed to date by these endoscopists ranged from less than 1000 to more than 10 000, with those having performed 1000–5000 procedures found to be the largest group (33.0%), followed by those with 5000–10 000 (28.7%), and those with more than 10 000 procedures (20.2%). More than 30% of the endoscopists were affiliated with Osaka University School of Medicine, 20% with Hyogo College of Medicine, and 11.6% with Osaka Medical College. It was found that about 37% of the endoscopists perform an average of 6–10, 25% perform 11–20, 20% perform up to five, and the rest perform more than 20 endoscopic procedures on a weekly basis. Eighty percent of the endoscopists showed an interest in reflux esophagitis. Only one-third of the endoscopists used M esophagitis as a grading, which might possibly have a negative impact on the study results. A total of 92.6% of them often made a diagnosis of hiatal hernia, and about 60% identified the presence of palisade vessels as the diagnostic hallmark of Barrett's epithelium, with 35.5% identifying the upper end of the gastric folds as the hallmark symptom (Table 2).

Assessment of video clips

Overall, 33 video clips were evaluated and rated by the 96 endoscopists, resulting in a total of 3168 ratings. Up to 3% (99) of the ratings were missing, with the responders ranging from 93 to 96 per video clip.

Interobserver agreement among endoscopists

The diagnoses of esophagitis, hiatal hernia, and Barrett's epithelium varied markedly among the endoscopists (Fig. 1). The κ -values in Table 3 show the level of agreement among all endoscopists for each diagnosis, that is, grades N, M, and A esophagitis, and for the diagnoses of hiatal hernia and Barrett's epithelium, for all video clips evaluated (not including duplicated videos).

The κ -value for all 91 endoscopists was 0.094, with a standard error of 0.002, indicating poor interobserver agreement in this group using this diagnostic method. Endoscopists showed the best agreement when diagnosing grade A esophagitis (0.167), indicating poor agreement, and the lowest κ -value for the diagnosis of grade M esophagitis (0.033). Kendall's coefficient of concordance, which assessed interobserver agreement considering the order of grades N, M, and A was 0.218, indicating a weak agreement. For the diagnoses of hiatal hernia and Barrett's epithelium, agreement between the endoscopists was slightly better, with κ -values of 0.177 and 0.159, respectively, both of which again indicate poor agreement among endoscopists.

Table 2 Participant characteristics

		Proportion (n = 96)*
Age, years	20–29	4.3%
	30–39	33.7%
	40–49	52.2%
	50–59	8.7%
	60+	1.1%
Sex	Men	92.6%
	Women	7.4%
Working style	University physician	13.7%
	Hospital physician	66.3%
	General practitioner	20.0%
Specialty	Internal medicine	80.0%
	Surgery	15.8%
	Other	4.2%
No. of years with EE	< 5 years	10.5%
	5–10 years	24.2%
	11–20 years	46.3%
	> 20 years	18.9%
No. of EE performed to date	< 1000	18.1%
	1000–5000	33.0%
	5001–10 000	28.7%
	> 10 000	20.2%
Institution trained at	Osaka University	31.6%
	Hyogo College of Medicine	20.0%
	Osaka Medical College	11.6%
	Kyoto University	2.1%
	Kobe University	1.1%
	Kansai Medical School	1.1%
	Kinki University	1.1%
	Osaka University/Kinki University/Other	1.1%
	Other	30.5%
Current weekly EE workload	≤ 5	20.0%
	6–10	36.8%
	11–20	25.3%
	20+	17.9%
Interest in RE	Yes	80.0%
	No	20.0%
Often use grade M	Yes	34.7%
	No	65.3%
Often diagnose hiatal hernia	Yes	92.6%
	No	7.4%
Hallmark of Barrett's epithelium	Palisade vessels	58.1%
	Upper end of gastric folds (contralateral)	35.5%
	Both of above	6.5%

*There were 1–4 missing value(s) for each variable. EE, endoscopic examination; RE, reflux esophagitis.

Table 3 The interobserver agreement of the diagnoses of grade N, M, A esophagitis, hiatal hernia, and Barrett's epithelium

		Grade of esophagitis (n = 91)				Hiatal hernia	Barrett's epithelium
		N	M	A	All		
Kappa	0.094	0.113	0.033	0.167	0.094	0.177	0.159
SE of kappa	0.002	0.003	0.003	0.003	0.002	0.004	0.004
Kendall*	0.218	–	–	–	0.218	–	–

Since the video clips of nos. 4 and 25, 5 and 29, and 8 and 32 were identical video pairs, nos. 25, 29, and 32 were excluded from calculation of the kappa values.

*Kendall's coefficient of concordance for ordinal response.

Reproducibility of the endoscopic diagnosis of grades N, M, and A esophagitis

The diagnoses of grades N, M, and A esophagitis, as well as hiatal hernia and Barrett's epithelium, made by the endoscopists were evaluated for their reproducibility by an analysis of symmetry of

assessments of the paired identical video clips #4/#25, #5/#29, and #8/#32. The κ -values in Tables 4 and 5 show the level of reproducibility among all endoscopists for each diagnosis, that is, grades N, M, and A esophagitis, as well as hiatal hernia and Barrett's epithelium, for the paired identical video clips.

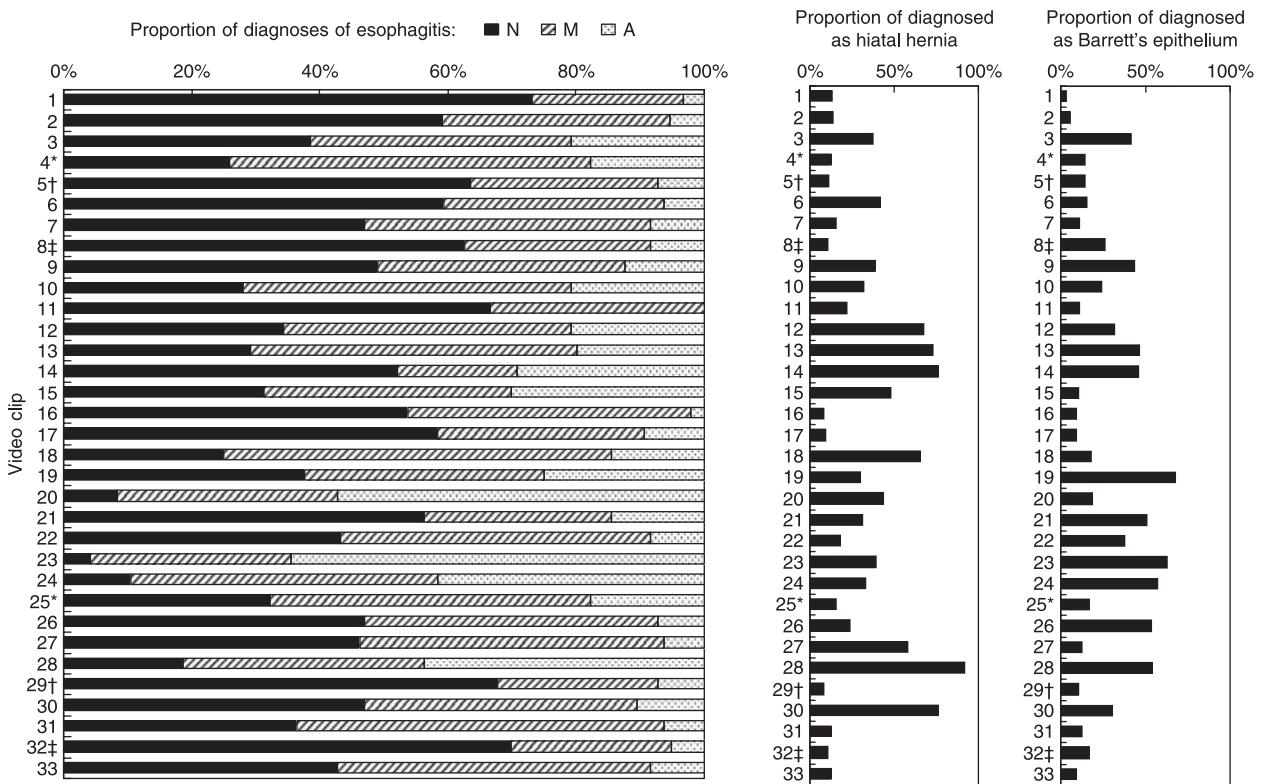


Fig. 1 Proportions of diagnoses of esophagitis (grades N, M, and A), hiatal hernia, and Barrett's epithelium among Japanese endoscopists. The number of endoscopists ranged from 93 to 96. Clips 4 and 25, 5 and 29, and 8 and 32 were identical video pairs used for assessing intraobserver reproducibility.

The κ -values for the diagnoses of grades N, M, and A esophagitis on video clips #4/#25, #5/#29, and #8/#32 were 0.275, 0.301, and 0.315, respectively, with a standard error of 0.083, 0.090, and 0.091, indicating fair intraobserver reproducibility among the endoscopists. The κ -values for the diagnoses of hiatal hernia on video clips #4/#25, #5/#29, and #8/#32 were 0.211, 0.430, and 0.235,

respectively, with a standard error of 0.124, 0.136, and 0.110, indicating fair agreement, while the κ -values for the diagnoses of Barrett's epithelium on the same video pairs were slightly better at 0.354, 0.476, and 0.442, respectively, with a standard error of 0.132, 0.149, and 0.148, indicating fair to moderate intraobserver reproducibility among the endoscopists.

Table 4 Reproducibility of the diagnoses of grade N, M, and A esophagitis

Grade	N	M	A	All	Kappa	SE
Video #4\#25						
N	14	9	2	25	0.297	0.104
M	15	32	7	54	0.355	0.099
A	2	7	8	17	0.330	0.099
All	31	48	17	96	0.275	0.083
Video #5\#9						
N	49	10	2	61	0.208	0.099
M	15	11	2	28	0.211	0.107
A	1	3	3	7	0.316	0.106
All	65	24	7	96	0.301	0.090
Video #8\#32						
N	49	10	1	60	0.357	0.121
M	13	13	2	28	0.384	0.174
A	5	1	2	8	0.260	0.172
All	67	24	5	96	0.315	0.091

The values are number of raters unless otherwise indicated.

Nos. 4 and 25, 5 and 29, and 8 and 32 were identical video clip pairs. None of them was significant by Bowker's test of symmetry (i.e., there were no significant differences in the frequency of N, M, and A grades between the first and second readings).

Table 5 Reproducibility of the diagnoses of hiatal hernia and Barrett's epithelium

Diagnosis	Hiatal hernia				Barrett's epithelium			
	(+)	(-)	All	Kappa (SE)	(+)	(-)	All	Kappa (SE)
Video #4\#25								
(+)	5	9	14	0.211	6	6	12	0.354
(-)	11	71	82	(0.124)	9	74	83	(0.132)
All	16	80	96		15	80	95	
Video #5\#29								
(+)	6	8	14	0.430	5	6	11	0.476
(-)	4	77	81	(0.136)	3	82	85	(0.149)
All	10	85	95		8	88	96	
Video #8\#32								
(+)	8	17	25	0.235	5	5	10	0.442
(-)	8	63	71	(0.110)	5	81	86	(0.148)
All	16	80	96		10	86	96	

The values are number of raters unless otherwise indicated.

Nos. 4 and 25, 5 and 29, and 8 and 32 are identical pairs of video clips. None of them was significant by McNemar's test of symmetry (i.e., there were no significant differences in the frequency of hiatal hernia and Barrett's epithelium between the first and second readings).

Factors affecting κ -values for diagnoses of esophagitis, hiatal hernia, and Barrett's epithelium

Table 6 demonstrates the effects of clinical practice variables and relevant experience of the endoscopists on their agreement in diagnosing cases shown in the video clips.

The reproducibility, determined from the likelihood of making the same diagnosis on the identical video clip at a later viewing, was greatest for general practitioners (moderate; 0.415 for the grade of esophagitis, 0.460 for hiatal hernia, and 0.441 for Barrett's epithelium) compared to hospital or university hospital physicians (fair; 0.250, 0.239, and 0.403, respectively). The highest reproducibility, 0.601 (good), was found for the diagnosis of Barrett's epithelium by experienced endoscopists, performing more than 5000 endoscopic examinations. However, those with a lower endoscopic examination workload (≤ 10 per week) had a higher reproducibility (moderate; 0.482) for diagnosing Barrett's epithelium than those with a greater workload (fair; 0.272), suggesting that these doctors may have more time to evaluate endoscopic lesions.

The highest interobserver agreement (κ -value of 0.280; fair) was found for the diagnosis of grade A esophagitis by endoscopists claiming to not often diagnose hiatal hernia ($n = 6$), compared to 0.160 (poor) for those for whom hiatal hernia was a more common diagnosis ($n = 85$). However, these results should be interpreted carefully because of the small sample size in the former group. The next highest interobserver agreement, 0.264 (fair), was found for the diagnosis of the 15 easy video clips, as compared to 0.177 (poor), the κ -value for the 15 difficult video clips.

DISCUSSION

It is now well accepted that the LA classification system and the MUSE (metaplasia, ulcer, stricture, erosion) scoring system facilitate reliable interobserver agreement among endoscopists in the diagnosis of reflux esophagitis grade A or greater (erosive esophagitis).⁴ In Japan, many endoscopists use 'the modified LA system' that includes minimal changes as constituting a distinct grade of reflux esophagitis. In this classification, minimal change esophagitis is characterized by the mucosa such as erythema and/or whitish turbidity.^{1,6} These minimal changes were eventually excluded from the diagnostic criteria for the LA classification system, as their recognition could be difficult depending on the experience of the endoscopists.²

Of the published classifications, furthermore, this modified LA classification in Japan (1,6) is the only one featuring not only erythematous but also acanthotic changes as the hallmark of mild reflux esophagitis. Some classifications have actually included erythema as endoscopic findings as this change¹ is more easily recognized than acanthosis, thus allowing good agreement (κ -values ranging from 0.6 to 0.8),² but poor agreement has been reported regarding acanthotic changes (κ -values ranging from 0.0 to 0.1).³ Nevertheless, many Japanese endoscopists believe in the 'modified LA system', and actually use it in their clinical practice. In addition, patients with this change are treated as having reflux esophagitis because minimal change (grade M esophagitis) is classified as esophagitis. However, unfortunately, clinical and/or pathological validation has been rarely conducted regarding whether minimal change (grade M) esophagitis should be added as a distinct grade of reflux esophagitis. Accordingly, we investigated

Table 6 Factors that affect the kappa values for the diagnosis of grade N, M, A+ esophagitis, hiatal hernia, and Barrett's epithelium

Factor	No.	Grade of esophagitis				Reproducibility for all grades*	Hiatal hernia	Reproducibility*	Barrett's epithelium	Reproducibility*
		N	M	A	All					
Type of practice										
HP/UP	74	0.111	0.032	0.152	0.089	0.025	0.165	0.239	0.158	0.403
GP	18	0.111 (0.997)	0.051 (0.256)	0.227 (< 0.001)	0.118 (0.019)	0.415	0.213 (0.044)	0.460	0.150 (0.762)	0.441
Specialty										
Internists	73	0.108	0.032	0.155	0.088	0.333	0.164	0.312	0.152	0.420
Surgeons	15	0.118 (0.613)	0.035 (0.882)	0.236 (< 0.001)	0.111 (0.124)	NC	0.241 (0.005)	NC	0.215 (0.024)	NC
No. of years with EE										
≤ 10 years	33	0.062	0.004	0.124	0.054	0.373	0.179	0.285	0.151	0.259
≥ 11	59	0.139 (< 0.001)	0.048 (< 0.001)	0.188 (< 0.001)	0.114 (< 0.001)	0.254	0.168 (0.429)	0.300	0.160 (0.499)	0.431
No. of EE performed										
≤ 5000	47	0.087	0.009	0.163	0.076	0.293	0.198	0.275	0.156	0.253
≥ 5001	44	0.144 (< 0.001)	0.058 (> 0.001)	0.176 (0.120)	0.116 (> 0.001)	0.270	0.160 (0.002)	0.278	0.152 (0.795)	0.601
Current weekly EE workload										
≤ 10	51	0.104	0.019	0.161	0.085	0.271	0.178	0.306	0.134	0.482
≥ 11	41	0.115 (0.215)	0.042 (0.006)	0.166 (0.522)	0.097 (0.057)	0.329	0.167 (0.398)	0.298	0.185 (< 0.001)	0.272
Interested in RE?										
Yes	73	0.115	0.036	0.168	0.096	0.307	0.198	0.378	0.180	0.377
Not much	19	0.089 (0.070)	0.001 (0.014)	0.137 (0.032)	0.068 (0.009)	NC	0.088 (< 0.001)	NC	0.082 (< 0.001)	NC
Often use grade M?										
Yes	32	0.098	0.039	0.172	0.089	0.321	0.155	0.317	0.163	0.390
Not often	60	0.122 (0.010)	0.030 (0.333)	0.157 (0.115)	0.097 (0.246)	0.245	0.176 (0.127)	0.278	0.155 (0.573)	0.415
Often diagnose HH?										
Yes	85	0.111	0.035	0.160	0.092	0.294	0.175	0.334	0.163	0.385
Not often	6	0.157 (0.332)	-0.006 (0.380)	0.280 (0.011)	0.128 (0.297)	NC	0.067 (0.105)	NC	0.223 (0.375)	NC
Hallmark of BE										
Palisade vessels	53	0.119	0.048	0.184	0.106	0.332	0.200	0.385	0.174	0.346
UEGF (contralateral)	31	0.091 (0.005)	0.011 (< 0.001)	0.147 (< 0.001)	0.071 (< 0.001)	0.278	0.154 (0.001)	0.169	0.133 (0.006)	0.454
Level of difficulty										
15 easy clips	93	0.147	0.048	0.199	0.122	-	0.264	-	0.186	-
15 difficult clips	93	0.078 (> 0.001)	0.017 (> 0.001)	0.119 (> 0.001)	0.062 (> 0.001)	-	0.177 (> 0.001)	-	0.159 (> 0.001)	-

Values are kappa-values unless otherwise indicated. *P*-values are for difference between two groups of each factor.

*Mean of the 3 kappa-values for videos nos. 4 and 25, 5 and 29, and 8 and 32. NC, not calculable because of the small sample size; HP, community hospital physician; UP, university hospital physician; GP, general practitioner; EE, endoscopic examinations; RE, reflux esophagitis; HH, hiatal hernia; UEGF, upper end of the gastric folds; BE, Barrett's epithelium.

inter- and intraobserver agreement in the endoscopic diagnosis of grade M esophagitis as a first step in validating this classification.

One hundred Japanese endoscopists in the Hanshin area participated in this study. The Hanshin area is located between Kobe and Osaka, supporting an estimated 2.5 million people. As shown in Table 2, the participants had various backgrounds in terms of not only experience of endoscopic procedures but also working style, educational institution, interest in esophagitis, and whether they often use grade M esophagitis. Although the majority of participants did not routinely use grade M this time, our results could be generalized to ordinary Japanese endoscopists.

Furthermore, besides the investigation of interobserver diagnostic agreement, duplicate video clips were used to examine intraobserver agreement. In other words, the reproducibility of the diagnoses was examined to reveal the difficulty of diagnosis only by color change. In the current study, we used short video clips, not still photographs, because we thought it may more closely reflect the daily clinical setting. In daily practice, relatively quick endoscopic evaluation is done, and an individual endoscopist makes a diagnosis based on their impression of different parts of moving images. In contrast, still photographs provide a single image that would limit diagnostic observation, leading to rather similar evaluations compared with video clips. However, the usage of short video clips to assess interobserver variation in endoscopic diagnosis has limitations. In performing GI endoscopy, we usually perform repositioning of the endoscope and reexamination of the lesions, which may be very important for making a diagnosis. However, video clips do not allow us to do this routine procedure, possibly lowering the diagnostic power of the examination.

The current study consistently indicated poor agreement regarding the diagnoses of minimal change esophagitis, as well as the fair reproducibility of these diagnoses among participating endoscopists, regardless of their age, type of practice, past endoscopic experience, or current endoscopic workload. An important point is that this poor agreement may not have been due to the video clips having been made only from very difficult cases. We made video clips from endoscopic examinations of randomly selected patients with normal to mild reflux esophagitis on a particular day. However, it would be technically difficult to externalize whether the diagnosis of a particular case is easy or difficult. Therefore, in this study, the level of difficulty in diagnosis was also assessed by the use of a visual analog scale for each presented case. In comparison of the 15 easy and difficult cases, the kappa values of these cases were 0.122 and 0.062, respectively, and the easy cases showed a significantly better value than that of the difficult cases. However, even for easy cases, the kappa value

was 0.122, still suggesting a limitation in diagnosing grade M esophagitis.

The primary and most important factor for the generally poor diagnostic agreement using this classification is that the diagnostic criteria are based on subjective changes in color or discoloration. In other words, the definition of grade M is very vague. Another factor seems to lie in the lack of education regarding this classification. Namely, endoscopists neither have opportunities to learn the diagnostic skill nor to exchange diagnostic information. It may also be partially due to the fact that endoscopic findings of mild esophagitis vary rapidly, and a definite pathological diagnosis is difficult as biopsy or surgical operation is rarely performed for this type of esophagitis.

Based on the results obtained from this study regarding inter and/or intraobserver agreement of grade M esophagitis, the modified LA system should not be used in clinical practice. The LA classification system may allow the majority of major findings to be detected, and thus be adequate for diagnostic purposes. Our study results also suggest that endoscopic diagnosis of reflux esophagitis may not be so accurate in daily practice, because the kappa value for not only the diagnosis of grade M but also that of grade A is not satisfactory. However, in the daily clinical setting, there may not be too much point in being overly specific about the diagnosis. Generally, endoscopic diagnosis does not provide definitive clues regarding the treatment strategy or predict the outcome of reflux esophagitis.¹²⁻¹⁴ Over the last 30 years, clinical studies have focused primarily on the management of patients with erosive esophagitis. However, in recent years NERD has become the focus of attention,^{15,16} and symptoms have been regarded as the primary target of the treatment in NERD and/or mild erosive reflux.^{16,17}

However, grade M reflux esophagitis may suggest an early form of reflux esophagitis, particularly if acanthotic or whitish changes are observed. Such changes are regarded as indicating a thickening of the epithelia due to exposure to acid regurgitation.¹⁸ Further research on the pathological basis of grade M esophagitis would possibly help identify nonerosive esophagitis with excess exposure to acid regurgitation.

References

- 1 Hongo M. Minimal changes in reflux esophagitis: red ones and white ones. *J Gastroenterol* 2006; 41: 95-9.
- 2 Armstrong D, Bennett J R, Blum A L *et al*. The endoscopic assessment of esophagitis: a progress report on observer agreement. *Gastroenterology* 1996; 111: 85-92.
- 3 Lundell L R, Dent J, Bennett J R *et al*. Endoscopic assessment of oesophagitis: clinical and functional correlates and further validation of the Los Angeles classification. *Gut* 1999; 45: 172-80.
- 4 Rath H C, Timmer A, Kunkel C *et al*. Comparison of interobserver agreement for different scoring systems for reflux esophagitis: Impact of level of experience. *Gastrointest Endosc* 2004; 60: 44-9.

- 5 Pandolfino J E, Vakil N B, Kahrilas P J. Comparison of inter- and intraobserver consistency for grading of esophagitis by expert and trainee endoscopists. *Gastrointest Endosc* 2002; 56: 639–43.
- 6 Hoshihara Y, Hashimoto M. Endoscopic classification of reflux esophagitis (in Japanese with English abstract). *Nippon Rinsho* 2000; 58: 1808–12.
- 7 Ismail-Beigi F, Horton R F, Pope C E. Histological consequences of gastroesophageal reflux in man. *Gastroenterology* 1970; 58: 163–74.
- 8 Ismail-Beigi F, Pope C E. Distribution of the histological changes of gastroesophageal reflux disease in the distal esophagus of man. *Gastroenterology* 1974; 66: 1109–13.
- 9 Takubo K, Honma N, Arya G *et al.* Is there a set of histologic changes that are invariably reflux associated? *Arch Pathol Laboratory Med* 2005; 129: 159–63.
- 10 Amano Y, Ishimura N, Furuta K *et al.* Interobserver agreement on classifying endoscopic diagnoses of nonerosive esophagitis. *Endoscopy* 2006; 38: 1032–5.
- 11 Fleiss J L. *Statistical Methods for Rates and Properties*, 3rd edn. New York: John Wiley & Sons, Inc., 2003.
- 12 Moayyedi P, Talley N J. Gastro-oesophageal reflux disease. *Lancet* 2006; 367: 2086–100.
- 13 Armstrong D, Marshall J K, Chiba N *et al.* Canadian Association of Gastroenterology GERD Consensus Group. Canadian Consensus Conference on the management of gastroesophageal reflux disease in adults – update 2004. *Can J Gastroenterol* 2005; 19: 15–35.
- 14 DeVault K R, Castell D O; American College of Gastroenterology. Updated guidelines for the diagnosis and treatment of gastroesophageal reflux disease. *Am J Gastroenterol* 2005; 100: 190–200.
- 15 Labenz J, Malfertheiner P. Treatment of uncomplicated reflux disease. *World J Gastroenterol* 2005; 11: 4291–9.
- 16 Fock K M, Talley N, Hunt R *et al.* Report of the Asia-Pacific consensus on the management of gastroesophageal reflux disease. *J Gastroenterol Hepatol* 2004; 19: 357–67.
- 17 Bytzer P. Goals of therapy and guidelines for treatment success in symptomatic gastroesophageal reflux disease patients. *Am J Gastroenterol* 2003; 98: S31–9.
- 18 Kiesslich R, Kanzler S, Vieth M *et al.* Minimal change esophagitis: prospective comparison of endoscopic and histological markers between patients with non-erosive reflux disease and normal controls using magnifying endoscopy. *Dig Dis* 2004; 22: 221–7.