Evaluating the Endoscopic Reference Score for eosinophilic esophagitis: moderate to substantial intra- and interobserver reliability

Authors Bram D. van Rhijn, Marijn J. Warners, Wouter L. Curvers, Anja U. van Lent, Noor L. Bekkali, R. Bart Takkenberg, Jaap J. Kloek, Jacques J. G. H. M. Bergman, Paul Fockens, Albert J. Bredenoord

Institution Department of Gastroenterology and Hepatology, Academic Medical Center, University of Amsterdam, Amsterdam, The Netherlands

Corresponding author
Bram D. van Rhijn, MD
Department of
Gastroenterology
and Hepatology
Academic Medical Center,
Amsterdam
PO Box 22660
1100 DD Amsterdam
The Netherlands
Fax: +31-20-6917033
b.d.vanrhijn@amc.uva.nl

Background and study aims: Recently the Endoscopic Reference Score (EREFS) for endoscopic assessment of eosinophilic esophagitis was introduced, with good interobserver agreement for most signs. The EREFS has not yet been evaluated by other investigators and intraobserver agreement has not been assessed. The aim of this study was to further validate the EREFS by assessing interobserver and intraobserver agreement of endoscopic signs in patients with eosinophilic esophagitis.

Patients and methods: High-quality endoscopic images were made of the esophagus of 30 patients with eosinophilic esophagitis (age 36 years, range 23 – 46 years; 5 female), 6 of whom were in remission. At least three depersonalized images per patient were incorporated into a slideshow. Images were scored by four expert and four trainee endoscopists who were blinded to the patients' conditions. Interobserver agreement was assessed.
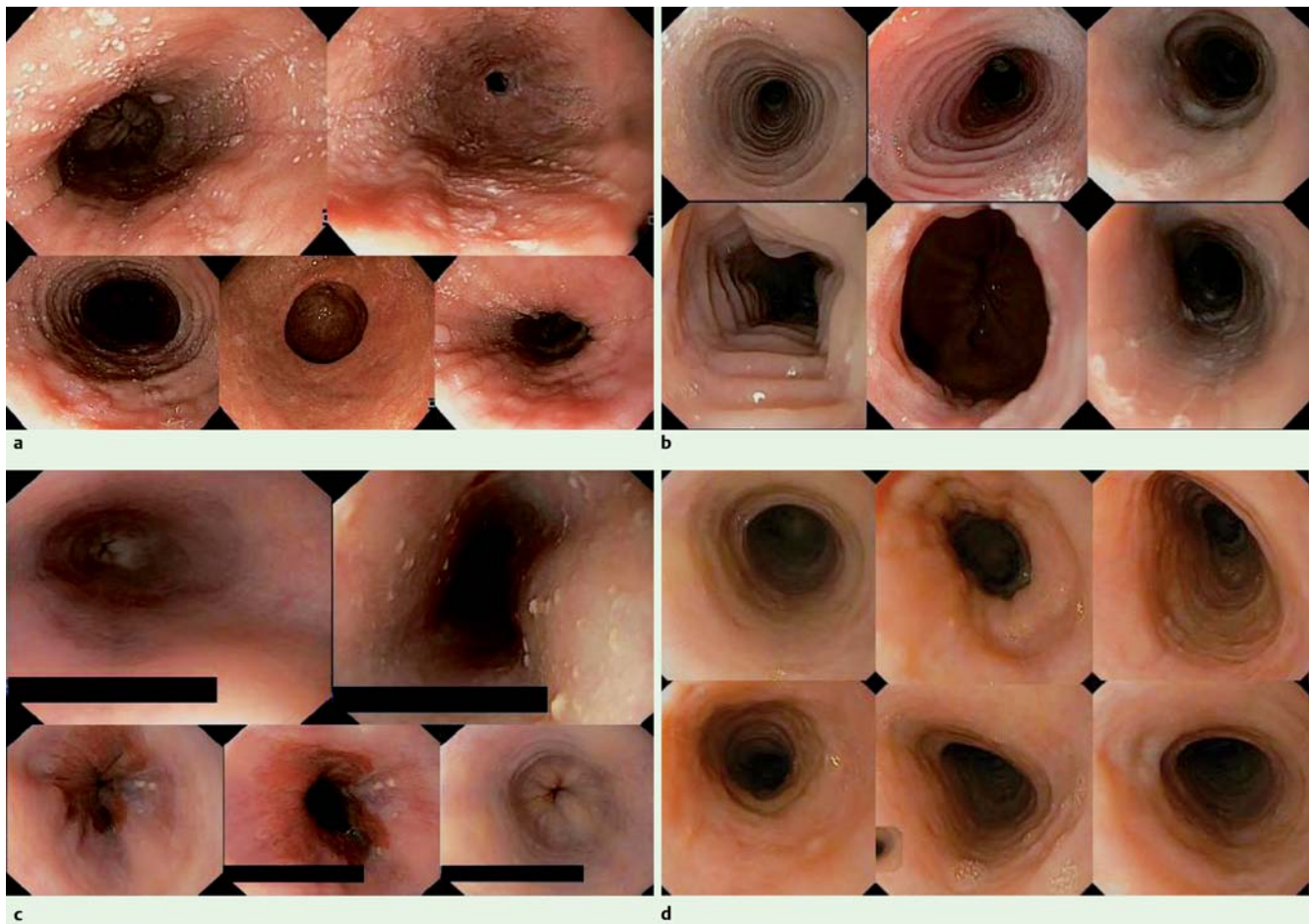
After 4 weeks, the images were rescored in a different order to assess intraobserver agreement.
Results: Interobserver agreement was substantial for rings (κ 0.70), white exudates (κ 0.63), and crepe paper esophagus (κ 0.62), moderate for furrows (κ 0.49) and strictures (κ 0.54), and slight for edema (κ 0.12). Intraobserver agreement was substantial for rings (median κ 0.64, IQR 0.46 – 0.70), furrows (median κ 0.69, IQR 0.50 – 0.89), and crepe paper esophagus (median κ 0.69, IQR 0.62 – 0.83), moderate for white exudates (median κ 0.58, IQR 0.54 – 0.71) and strictures (median κ 0.54, IQR 0.33 – 0.70), and less than chance for edema (median κ 0.00, IQR 0.00 – 0.29). Inter- and intraobserver agreement was not substantially different between expert and trainee endoscopists.
Conclusions: Using the EREFS, endoscopic signs of eosinophilic esophagitis were scored consistently by expert and trainee endoscopists.

Introduction
▼

Eosinophilic esophagitis is an increasingly diagnosed chronic inflammatory disease of the esophagus, causing symptoms of dysphagia and food impaction [1,2]. Endoscopically, concentric rings, linear furrows, white exudates, edema, strictures, diffuse narrowing, and crepe paper appearance can be identified, although in some patients none of these signs are present [3 – 5]. Endoscopic signs of eosinophilic esophagitis are not pathognomonic for eosinophilic esophagitis, and are scored with variable reliability [5,6]. Therefore, the diagnosis of eosinophilic esophagitis cannot be based solely on endoscopic signs, and requires esophageal biopsy sampling [7]. Recently, a validated classification system for the endoscopic assessment of esophageal signs of eosinophilic esophagitis (EREFS) has been intro-

duced, showing good interobserver agreement for most signs [8]. The EREFS enables uniform scoring of endoscopic signs. It has not yet been evaluated by other investigators and the intraobserver agreement has not been assessed. The value of the EREFS is thus not completely known. If the good interobserver agreement could be confirmed, and if intraobserver agreement also appears to be good, this would endorse application of the classification system in clinical trials and in the follow-up of patients with eosinophilic esophagitis. The aim of the current study, therefore, was to further validate the classification system by assessing interobserver and intraobserver agreement of endoscopic signs in patients with eosinophilic esophagitis, both in active disease and in remission.

Fig. 1 Examples of endoscopic images of eosinophilic esophagitis that were scored by the observers. **a** Active disease, 110 eosinophils per high-power field (eos/hpf); **b** active disease, 65 eos/hpf; **c** remission, 0 eos/hpf; **d** remission, 6 eos/hpf.

Patients and methods
▼

This prospective observational study was conducted at the Academic Medical Center, Amsterdam, The Netherlands, which functions as a tertiary referral center for patients with eosinophilic esophagitis. Patients were not exposed to any additional interventions for the study purpose, and therefore the need for formal medical ethical assessment was waived by the institutional review board, in accordance with Dutch law (reference number W13_134#13.17.0172).

Data collection

A total of 30 patients who were diagnosed with eosinophilic esophagitis according to current guidelines (symptoms of esophageal dysfunction and histology showing >15 eosinophils per high-power field [eos/hpf] despite double-dose proton pump inhibitor treatment) [7] were included in the study. Six of the patients were in remission during endoscopy (<15 eos/hpf), and were included in this study as a means to avoid observer bias. Prospectively collected endoscopic images were selected based on quality (i.e. sharpness and lack of motion effects). For each patient, 3 – 6 depersonalized images were selected. Each of the 30 slide sets thus contained a selection of images from one endoscopy in an individual patient. Observers were aware of the study design, but did not know how many patients were in remission.

Observers

Four expert endoscopists and four trainees scored the images. Expert endoscopists were consultant gastroenterologists who had performed upper endoscopy in at least 40 patients with eosinophilic esophagitis. Trainees were all residents in Gastroenterology who had at least 2 years of endoscopy experience and had performed endoscopy in five or fewer eosinophilic esophagitis patients.

Image evaluation

Each observer read the paper describing the EREFS for endoscopic signs of eosinophilic esophagitis before evaluation of the endoscopic images [8]. The observers also viewed an introductory slideshow (Microsoft PowerPoint 2003; Microsoft Inc., Redmond, Washington, USA) containing representative images and written descriptions of the EREFS. Subsequently, the observers, who were blinded to the patients' conditions, scored the endoscopic images of the patient slide sets according to the EREFS (**◕ Fig. 1**) [8]. The interobserver agreement of EREFS scores for endoscopic signs was then assessed. After 4 weeks, the same images were rescored in a different order in order to assess the intraobserver agreement. Responses were entered onto a standardized printed form.

Table 1 Interobserver agreement of EREFS and alternative grading.

Endoscopic sign[1]	Pairwise agreement[2], n (%)	κ [95%CI]	Agreement interpretation
Rings	**492 (59)**	**0.70 [0.58 – 0.82][3]**	**Substantial**
Absent/mild collapsed	585 (70)	0.60 [0.46 – 0.74][3]	Substantial
Mild/moderate collapsed	626 (75)	0.67 [0.54 – 0.79][3]	Substantial
Moderate/severe collapsed	558 (66)	0.71 [0.59 – 0.82][3]	Substantial
Absent vs. present	744 (89)	0.64 [0.57 – 0.71][4]	Substantial
Absent/mild vs. moderate/severe	653 (78)	0.55 [0.48 – 0.61][4]	Moderate
White exudates	**572 (68)**	**0.63 [0.49 – 0.76][3]**	**Substantial**
Absent/mild collapsed	703 (84)	0.65 [0.58 – 0.72][4]	Substantial
Absent vs. present	680 (81)	0.42 [0.35 – 0.49][4]	Moderate
Furrows	**656 (78)**	**0.49 [0.42 – 0.56][4]**	**Moderate**
Absent/mild collapsed	637 (76)	0.32 [0.25 – 0.38][4]	Fair
Absent-mild-severe	476 (57)	0.54 [0.39 – 0.70][3]	Moderate
Edema	**751 (89)**	**0.12 [0.05 – 0.19][4]**	**Slight**
Absent/mild collapsed	524 (62)	0.24 [0.17 – 0.31][4]	Fair
Absent-mild-severe	455 (54)	0.32 [0.18 – 0.49][3]	Fair
Stricture	**648 (77)**	**0.54 [0.47 – 0.60][4]**	**Moderate**
Crepe paper esophagus	**765 (91)**	**0.62 [0.56 – 0.69][4]**	**Substantial**

CI, confidence interval.
[1] Grading according to current EREFS is marked in bold. Nonbold entries represent an alternative grading system.
[2] Based on 840 possible pairwise comparisons.
[3] Intraclass correlation coefficient.
[4] Multi-rater kappa.

Statistical analysis

Descriptive statistics were used to characterize the study population. Continuous data were described as median with interquartile range (IQR) unless otherwise described. IBM SPSS Statistics (version 20.0; IBM Corp., Armonk, New York, USA) was used for statistical analyses.

The interobserver agreement was estimated in the first scoring session with multi-rater kappa (κ) for dichotomous data, and the intraclass correlation coefficient (ICC) for ordinal data [9,10]. The ICC with 95% confidence interval (CI) was calculated using the two-way random model for absolute agreement. In addition, the percentage of pairwise agreement was calculated for all data [11]. For all endoscopic signs, the proportion of pairwise agreement results from a comparison of the grading of the endoscopic sign between each observer and the other observers. The percentage of all pairwise comparisons was reported where the grading of the endoscopic sign was identical for the two observers. Between eight observers, 28 pairwise comparisons were made per sign per slide (in total, 840 comparisons per sign). When separating results for experts and trainees (four observers in each group), 6 comparisons were made per sign per slide (in total, 180 comparisons per sign). The interobserver agreement was interpreted based on κ/ICC estimates.

The intraobserver agreement was calculated per observer by using Cohen's κ for the agreement between the first and second scoring sessions [12]. Bootstrapping was performed using 50 samples to obtain 95th percentile CI. Interpretation of the intraobserver agreement was based on Cohen's κ.

A κ of ≤0.00 was interpreted as less than chance agreement, 0.01 – 0.20 as slight agreement, 0.21 – 0.40 as fair agreement, 0.41 – 0.60 as moderate agreement, 0.61 – 0.80 as substantial agreement, and 0.81 – 1.00 as almost perfect agreement [13].

Results

Patient characteristics

Patients were aged 36 years (IQR 23 – 46 years), and five (17%) were female. A total of 24 patients had active disease (peak eosinophil count 38 [IQR 20 – 100]/hpf), whereas 6 patients were in remission with fluticasone propionate (peak eosinophil count 2.5 [IQR 0 – 6]/hpf). According to ImmunoCAP ISAC serum IgE testing (Phadia AB, Uppsala, Sweden), 21 patients (70%) were sensitized against aero allergens, 17 patients (57%) were sensitized against food allergens, and 9 patients (30%) were not sensitized to any of the tested allergens [14].
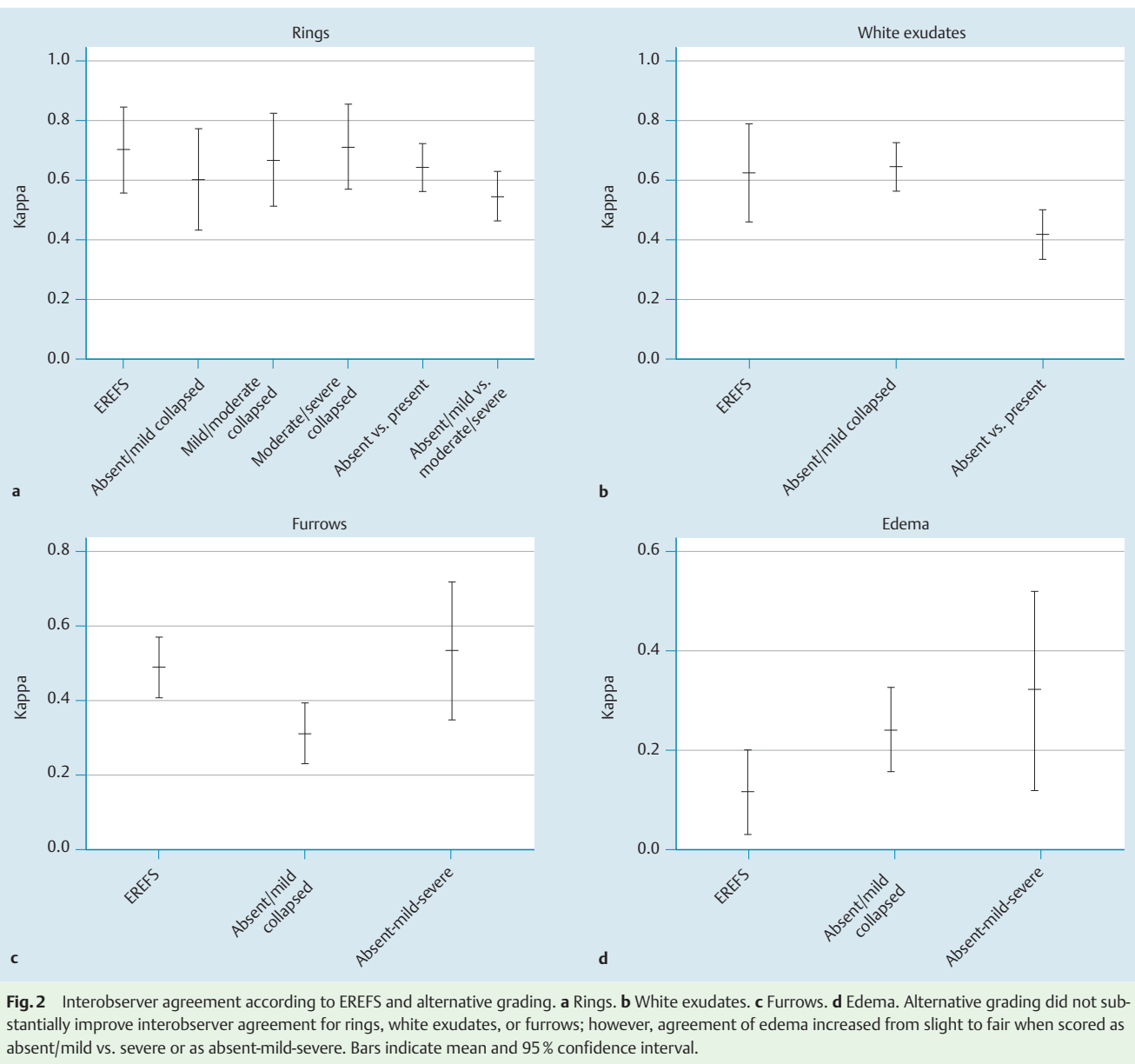
Endoscopic signs detected

In two scoring sessions by eight observers in 30 patients (480 observations), edema was detected most frequently (451 times, 94%), followed by rings (378 times, 79%), white exudates (366 times, 76%), furrows (331 times, 69%), strictures (218 times, 45%), and crepe paper esophagus (68 times, 14%). In each patient, at least one endoscopic sign of eosinophilic esophagitis was detected.

Interobserver agreement

When scored according to the EREFS, the interobserver agreement was substantial for rings, white exudates, and crepe paper esophagus, moderate for furrows and strictures, and slight for edema (⊙ **Table 1**). The interobserver agreement for edema increased to fair (κ 0.318, 95%CI 0.184 – 0.495) when categorized as absent-mild-severe (⊙ **Fig. 2**). Alternative grading did not substantially improve interobserver agreement for rings, white exudates, or furrows.

When comparing expert endoscopists with trainees, experts and trainees scored similar on rings (substantial), crepe paper esophagus (substantial), and strictures (moderate), and experts had slightly lower agreement for white exudates (moderate vs. substantial), furrows (moderate vs. substantial), and edema (less than chance vs. slight) (⊙ **Table 2**).

Fig. 2 Interobserver agreement according to EREFS and alternative grading. **a** Rings. **b** White exudates. **c** Furrows. **d** Edema. Alternative grading did not substantially improve interobserver agreement for rings, white exudates, or furrows; however, agreement of edema increased from slight to fair when scored as absent/mild vs. severe or as absent-mild-severe. Bars indicate mean and 95 % confidence interval.

Intraobserver agreement

When scored according to EREFS, the intraobserver agreement was substantial for rings (median κ 0.64, IQR 0.46 – 0.70), furrows (median κ 0.69, IQR 0.50 – 0.89), and crepe paper esophagus (median κ 0.69, IQR 0.62 – 0.83), moderate for white exudates (median κ 0.58, IQR 0.54 – 0.71) and strictures (median κ 0.54, IQR 0.33 – 0.70), and less than chance for edema (median κ 0.00, IQR 0.00 – 0.29). The intraobserver agreement for white exudates increased from moderate to substantial when scored as absent/mild vs. severe (median κ 0.69, IQR 0.61 – 0.83) and as absent vs. present (median κ 0.61, IQR 0.45 – 0.71) (○ **Fig. 3**). The intraobserver agreement for edema increased to moderate when scored as absent/mild vs. severe (median κ 0.49, IQR 0.29 – 0.70) and to fair when scored as absent-mild-severe (median κ 0.38, IQR 0.23 – 0.57).

When comparing expert endoscopists with trainees, experts had slightly higher intraobserver agreement for rings, white exudates, and strictures (all substantial vs. moderate), but slightly lower intraobserver agreement for furrows (moderate vs. sub-

stantial), crepe paper esophagus (substantial vs. almost perfect), and edema (less than chance vs. slight) (○ **Table 3**).

Discussion
▼

Using the recently developed classification system for endoscopic signs of eosinophilic esophagitis (EREFS), moderate to substantial interobserver agreement was demonstrated for rings, white exudates, strictures, and crepe paper esophagus. Furthermore, this study is the first to investigate intraobserver agreement using this novel endoscopic classification system. Moderate to substantial intraobserver agreement was found for rings, white exudates, furrows, strictures, and crepe paper esophagus.

The interobserver agreement for endoscopic signs in patients with eosinophilic esophagitis has not been frequently investigated [6,8]. In a prospective study using solitary, still images, Peery et al. showed moderate interobserver agreement for rings and furrows, whereas interobserver agreement for plaques (white

Table 2 Interobserver agreement of EREFS in expert and trainee endoscopists.

Endoscpic sign	Experts		Trainees	
	Pairwise agreement[1], n (%)	κ [95%CI]	Pairwise agreement[1], n (%)	κ [95%CI]
Rings	107 (59)	0.76 [0.62–0.86][2]	105 (58)	0.66 [0.50–0.80][2]
White exudates	112 (62)	0.53 [0.33–0.71][2]	131 (73)	0.69 [0.54–0.82][2]
Furrows	139 (77)	0.41 [0.26–0.55][3]	148 (82)	0.61 [0.46–0.75][3]
Edema	171 (95)	–0.03 [–0.17 to 0.12][3]	96 (53)	0.11 [–0.04 to 0.25][3]
Stricture	138 (77)	0.53 [0.38–0.68][3]	137 (76)	0.52 [0.37–0.66][3]
Crepe paper esophagus	162 (90)	0.61 [0.46–0.75][3]	167 (93)	0.67 [0.52–0.82][3]

CI, confidence interval.
[1] Based on 180 possible pairwise comparisons.
[2] Intraclass correlation coefficient.
[3] Multi-rater kappa.
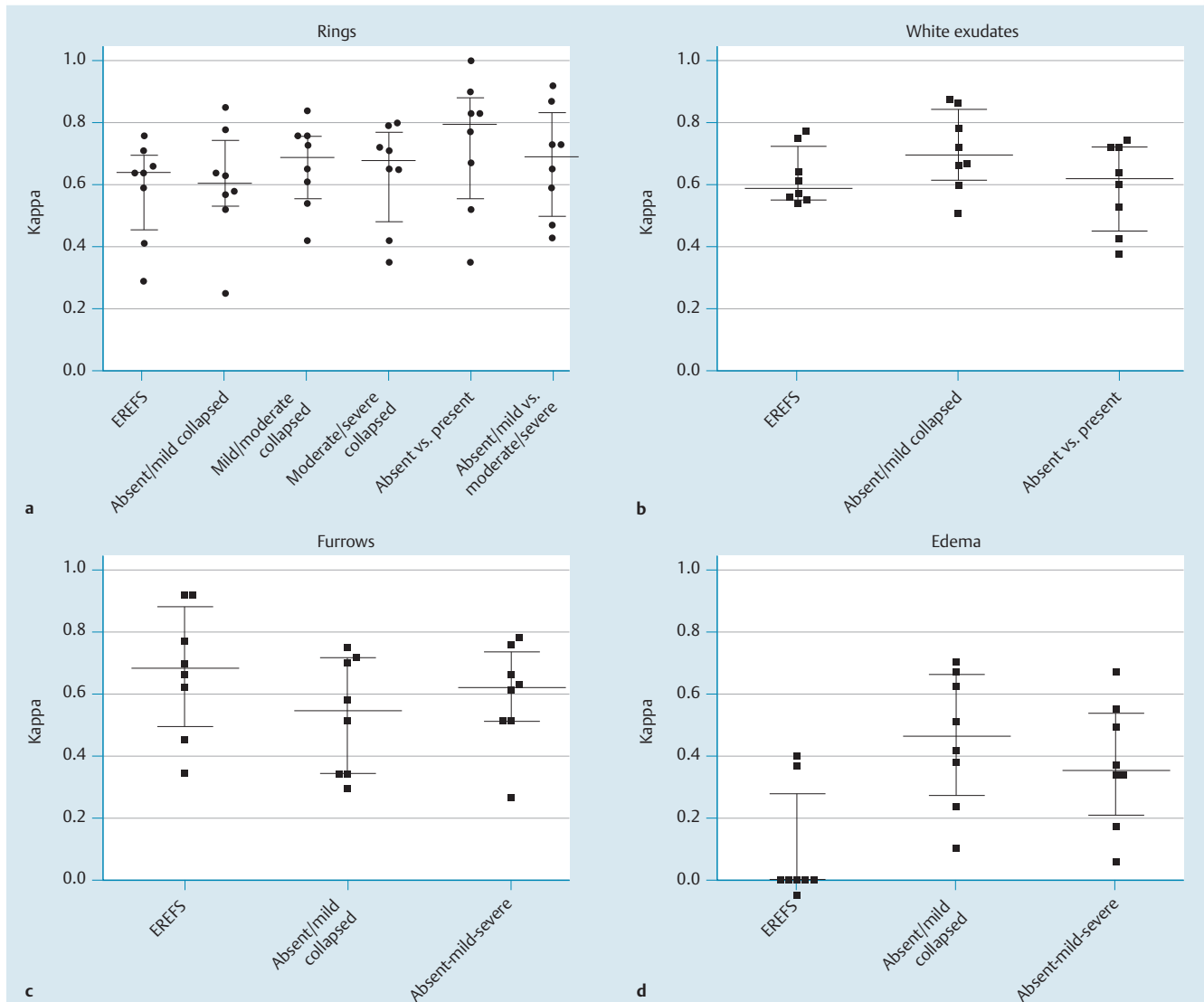
Fig. 3 Intraobserver agreement according to EREFS and alternative grading. **a** Rings. **b** White exudates. **c** Furrows. **d** Edema. Alternative grading did not substantially improve intraobserver agreement for rings or furrows; however, agreement could be increased for white exudates (from moderate to substantial when scored as absent/mild vs. severe) and for edema (from less than chance to moderate when scored as absent/mild vs. severe). Bars indicate median and interquartile range.

Table 3 Intraobserver agreement for EREFS in expert and trainee endoscopists.

Observer		Experts, κ [95%CI]				Trainees, κ [95%CI]				Total, median κ (range)	
endoscopic sign		I	II	III	IV	V	VI	VII	VIII	Experts	Trainees
Rings		0.64 [0.38–0.89]	0.66 [0.29–0.89]	0.76 [0.47–0.89]	0.71 [0.39–0.89]	0.41 [0.23–0.70]	0.64 [0.40–0.91]	0.29 [−0.03 to 0.51]	0.59 [0.26–0.77]	0.69 (0.64–0.76)	0.50 (0.29–0.64)
White exudates		0.63 [0.29–0.87]	0.60 [0.32–0.83]	0.53 [0.26–0.74]	0.76 [0.54–0.95]	0.55 [0.13–0.78]	0.74 [0.54–0.95]	0.56 [0.31–0.78]	0.54 [0.24–0.78]	0.62 (0.53–0.76)	0.56 (0.54–0.74)
Furrows		0.67 [0.32–0.93]	0.35 [−0.19 to 0.71]	0.93 [0.77–1.00]	0.46 [−0.08 to 1.00]	0.63 [0.14–0.90]	0.93 [0.70–1.00]	0.71 [0.43–0.96]	0.78 [0.54–1.00]	0.57 (0.35–0.93)	0.75 (0.63–0.93)
Edema		−0.05 [−0.10 to 0.00]	0.00 [0.00–0.00]	0.00 [0.00–0.00]	0.00 [0.00–0.00]	0.00 [0.00–0.00]	0.42 [−0.16 to 0.80]	0.00 [0.00–0.00]	0.39 [0.00–0.89]	0.00 (−0.05 to 0.00)	0.20 (0.00–0.42)
Stricture		0.80 [0.55–1.00]	0.54 [0.22–0.86]	0.28 [−0.09 to 0.70]	0.73 [0.42–0.93]	0.46 [0.01–0.86]	0.26 [−0.14 to 0.70]	0.59 [0.18–0.85]	0.53 [0.28–0.80]	0.64 (0.28–0.80)	0.50 (0.26–0.59)
Crepe paper esophagus		0.61 [−0.05 to 1.00]	0.67 [0.37–1.00]	0.71 [0.00–1.00]	0.63 [−0.05 to 1.00]	0.61 [−0.08 to 1.00]	0.90 [0.58–1.00]	0.78 [0.00–1.00]	0.84 [0.47–1.00]	0.65 (0.61–0.71)	0.81 (0.61–0.90)

CI, confidence interval.

Table 4 Adjusted EREFS for the endoscopic assessment of endoscopic signs of eosinophilic esophagitis, based on Hirano et al. [8].

Endoscopic sign	Classification
Rings	Grade 0: absent
	Grade 1: mild (subtle circumferential rings)
	Grade 2: moderate (distinct rings, still passage of diagnostic endoscope)
	Grade 3: severe (distinct rings, no passage of diagnostic endoscope)
White exudates	Grade 0: absent to mild (< 10 % of esophageal surface area)
	Grade 1: severe (> 10 % of esophageal surface area)
Furrows	Grade 0: absent
	Grade 1: present
Edema	Grade 0: absent to mild (loss of clarity of vascular markings)
	Grade 1: severe (absence of vascular markings)
Stricture	Grade 0: absent
	Grade 1: present
Crepe paper esophagus	Grade 0: absent
	Grade 1: present

exudates) was poor [6]. For rings and furrows, κ values in the current study were slightly higher than those in the Peery study, and agreement was also much higher for white exudates. Furthermore, the interobserver agreement obtained in the current study was very similar to agreement rates reported by Hirano et al. [8]. The intraobserver agreement for endoscopic signs in patients with eosinophilic esophagitis has been reported previously. In the study by Peery et al., intraobserver agreement for all endoscopic signs of eosinophilic esophagitis was variable [6]. In the current study, moderate to substantial intraobserver agreement was found for rings, white exudates, furrows, strictures, and crepe paper esophagus. Compared with the intraobserver agreement in Peery et al. [6] agreement for furrows was similar and that for rings and exudates was higher in the current study. This might reflect the fact that the Peery study was conducted before the EREFS was published; assessment of endoscopic signs was not standardized and endoscopic signs were not graded [8]. In the current study, the intraobserver agreement for white exudates could be increased from moderate to substantial, with persisting substantial interobserver agreement, when scored as absent/mild vs. severe. The interobserver and intraobserver agreement for edema also increased when scored as absent/mild vs. severe. Based on these findings, we propose that white exudates and edema be scored using alternative grading (○ Table 4).

There were no structural differences in agreement when comparing expert with trainee endoscopists. The results of the study by Hirano et al. demonstrated moderately better interobserver agreement for the assessment of white exudates, edema, and crepe paper esophagus by experts, whereas in the current study white exudates, edema, and furrows, and crepe paper esophagus were actually scored with slightly higher agreement by trainees than by experts [8]. Similarly, the intraobserver agreement for furrows, edema, and crepe paper esophagus were also higher for trainees in the current study. Experts had slightly higher intraobserver agreement for rings, white exudates, and strictures. However, differences were small, and we conclude that the EREFS can be used for reliable scoring, not only by experts but also by trainee endoscopists.

The EREFS for endoscopic signs of eosinophilic esophagitis thus enables consistent and uniform scoring of endoscopic signs, which facilitates both communication between clinicians and adequate comparison between clinical studies performed in patients with eosinophilic esophagitis. Although endoscopy has an established role in the diagnosis of eosinophilic esophagitis, the value of endoscopic signs in the follow-up of patients is still unclear. The current study confirms the moderate to substantial interobserver agreement rates described by Hirano et al., and adds valuable information with regard to the intraobserver agreement. The intraobserver agreement of the EREFS was moderate to substantial for most signs of eosinophilic esophagitis, which means that these signs were scored consistently. This is important for the reliable interpretation of endoscopic signs in clinical trials and in the follow-up of patients with eosinophilic esophagitis.

In line with the study by Hirano et al., [8], interobserver agreement was reported as κ estimates and pairwise agreement. Occasionally low κ estimates were found while pairwise agreement was high; this is a paradox that has been described in the literature [11]. For example, the widespread presence of edema even in patients in remission resulted in low κ estimates among observers, despite the high pairwise agreement between them. The observers thus scored edema reliably, although this was not reflected by the κ statistic. This is caused by the widespread presence of edema, even in patients in remission, which increases the a priori chance of detecting edema making it a less discriminatory sign [11]. Therefore, agreement based on a combination of the κ estimate and the percentage of agreement more comprehensively describes the agreement than the κ estimate alone.

This study has some limitations. First, the image set of 30 patients may seem small; however, similar studies have used comparable numbers of patients [6,8]. One of the factors that could influence agreement results is observer fatigue; scoring of 30 image sets already takes about 1 – 1.5 hours. Therefore, in order to avoid observer fatigue, we believe that 30 image sets is a reasonable number. Second, still images were used rather than endoscopy, which may hamper interpretation. For instance, endoscopy provides the opportunity to closely inspect a region that may not look clear at first sight. In addition, endoscopy offers the possibility to flush away white plaques that might resemble white exudates. To minimize the disadvantage of using still images, a set of 3 – 6 (median 5) images per patient were used, which incorporated different parts of the esophagus after flushing and cleansing of the esophagus in order to mimic real endoscopy. By using this set of still images, agreement was very comparable to the agreement reported by Hirano et al., in which videos were used [8].

This study has demonstrated that the recently proposed EREFS for the assessment of endoscopic signs of eosinophilic esophagitis is a reliable and stable measure that can be employed by expert as well as trainee endoscopists. This is important for the interpretation of endoscopic signs of eosinophilic esophagitis in clinical trials and during follow-up of patients. To improve inter- and in-traobserver agreement, we propose an alternative scoring system that contains slight changes to the grading of white exudates and edema.

In conclusion, we have demonstrated moderate to substantial interobserver and intraobserver agreement for most endoscopic signs of eosinophilic esophagitis, supporting the widespread use of the EREFS.

Acknowledgments
▼

References

1 *Liacouras CA, Furuta GT, Hirano I* et al. Eosinophilic esophagitis: updated consensus recommendations for children and adults. J Allergy Clin Immunol 2011; 128: 3 – 20
2 *van Rhijn BD, Verheij J, Smout AJ* et al. Rapidly increasing incidence of eosinophilic esophagitis in a large cohort. Neurogastroenterol Motil 2013; 25: 47 – 52
3 *Prasad GA, Talley NJ, Romero Y* et al. Prevalence and predictive factors of eosinophilic esophagitis in patients presenting with dysphagia: a prospective study. Am J Gastroenterol 2007; 102: 2627 – 2632
4 *Mackenzie SH, Go M, Chadwick B* et al. Eosinophilic oesophagitis in patients presenting with dysphagia – a prospective analysis. Aliment Pharmacol Ther 2008; 28: 1140 – 1146
5 *Kim HP, Vance RB, Shaheen NJ* et al. The prevalence and diagnostic utility of endoscopic features of eosinophilic esophagitis: a meta-analysis. Clin Gastroenterol Hepatol 2012; 10: 988 – 996
6 *Peery AF, Cao H, Dominik R* et al. Variable reliability of endoscopic findings with white-light and narrow-band imaging for patients with suspected eosinophilic esophagitis. Clin Gastroenterol Hepatol 2011; 9: 475 – 480
7 *Dellon ES, Gonsalves N, Hirano I* et al. ACG clinical guideline: Evidenced based approach to the diagnosis and management of esophageal eosinophilia and eosinophilic esophagitis (EoE). Am J Gastroenterol 2013; 108: 679 – 692
8 *Hirano I, Moy N, Heckman MG* et al. Endoscopic assessment of the oesophageal features of eosinophilic oesophagitis: validation of a novel classification and grading system. Gut 2013; 62: 489 – 495
9 *Fleiss JL, Nee JCM, Landis J.* Large sample variance of kappa in the case of different sets of raters. Psychol Bull 1979; 1979: 974 – 977
10 *Shrout PE, Fleiss JL.* Intraclass correlations: uses in assessing rater reliability. Psychol Bull 1979; 86: 420 – 428
11 *Feinstein AR, Cicchetti DV.* High agreement but low kappa: I. The problems of two paradoxes. J Clin Epidemiol 1990; 43: 543 – 549
12 *Cohen J.* A coefficient of agreement for nominal scales. Educational and Psychological Measurement 1960; 1960: 37 – 46
13 *Landis JR, Koch GG.* The measurement of observer agreement for categorical data. Biometrics 1977; 33: 159 – 174
14 *van Rhijn BD, van Ree R, Versteeg SA* et al. Birch pollen sensitization with cross-reactivity to food allergens predominates in adults with eosinophilic esophagitis. Allergy 2013; 68: 1475 – 1481