Clinical Gastroenterology and Hepatology 2019;17:1477–1488

ALIMENTARY TRACT

Variation in Endoscopic Activity Assessment and Endoscopy Score Validation in Adults With Eosinophilic Esophagitis

Alain M. Schoepfer,[*,a] Ikuo Hirano,[‡,a] Michael Coslovsky,[§] Marie C. Roumet,[§] Marcel Zwahlen,[§] Claudia E. Kuehni,[§] David Hafner,[§] Jeffrey A. Alexander,[‖] Evan S. Dellon,[¶] Nirmala Gonsalves,[‡] John Leung,[#] Christian Bussmann,[**] Margaret H. Collins,[‡‡] Robert O. Newbury,[§§] Thomas C. Smyrk,[‖‖] John T. Woosley,[¶¶] Guang-Yu Yang,[##] Yvonne Romero,[‖,***,‡‡‡] David A. Katzka,[‖] Glenn T. Furuta,[§§§] Sandeep K. Gupta,[‖‖‖] Seema S. Aceves,[¶¶¶] Mirna Chehade,[###] Jonathan M. Spergel,[****] Gary W. Falk,[‡‡‡‡] Brian A. Meltzer,[§§§§] Gail M. Comer,[‖‖‖‖] Alex Straumann,[¶¶¶¶] and Ekaterina Safroneeva,[§] on behalf of the International EEsAI Study Group

*Division of Gastroenterology and Hepatology, Centre Hospitalier Universitaire Vaudois/CHUV, Lausanne, Switzerland; ‡Division of Gastroenterology, Northwestern University Feinberg School of Medicine, Chicago, Illinois; §Institute of Social and Preventive Medicine, University of Bern, Bern, Switzerland; ‖Division of Gastroenterology and Hepatology, Mayo Clinic, Rochester, Minnesota; ¶Center for Esophageal Diseases and Swallowing, Division of Gastroenterology and Hepatology, University of North Carolina School of Medicine, Chapel Hill, North Carolina; #Center for Food Related Diseases, Division of Pediatric Allergy and Immunology, Division of Gastroenterology, Tufts Medical Center and Floating Hospital for Children, Boston, Massachusetts; **Viollier AG, Institute for Pathology, Basel, Switzerland; ‡‡Division of Pathology and Laboratory Medicine, Cincinnati Children's Hospital Medical Center, Cincinnati, Ohio; §§Department of Pathology, Rady Children's Hospital, University of California, San Diego, San Diego, California; ‖‖Department of Laboratory Medicine and Pathology, Mayo Clinic College of Medicine, Rochester, Minnesota; ¶¶Department of Pathology and Laboratory Medicine, University of North Carolina School of Medicine, Chapel Hill, North Carolina; ##Department of Pathology, Northwestern University Feinberg School of Medicine, Chicago, Illinois; ***Department of Otolaryngology, Mayo Clinic, Rochester, Minnesota; ‡‡‡GI Outcomes Unit, Mayo Clinic, Rochester, Minnesota; §§§Gastrointestinal Eosinophilic Diseases Program, Department of Pediatrics, University of Colorado School of Medicine; Digestive Health Institute, Children's Hospital Colorado, Aurora, Colorado; ‖‖‖Division of Pediatric Gastroenterology, Hepatology, and Nutrition, Riley Hospital for Children, Indiana University School of Medicine, Indianapolis, Indiana; ¶¶¶Division of Allergy and Immunology, Rady Children's Hospital, University of California, San Diego, San Diego, California; ###Departments of Pediatrics and Medicine, Mount Sinai Center for Eosinophilic Disorders, Jaffe Food Allergy Institute, Icahn School of Medicine at Mount Sinai, New York, New York; ****Divisions of Allergy and Immunology, Department of Pediatrics, The Children's Hospital of Philadelphia, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania; ‡‡‡‡Division of Gastroenterology, Department of Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania; §§§§Adare Pharmaceuticals, Inc, Lawrenceville, New Jersey; ‖‖‖‖Kimberton Drug Development Consulting, LLC, Phoenixville, Pennsylvania; and ¶¶¶¶Swiss EoE Clinic, Division of Gastroenterology, University Hospital Zurich, Zurich, Switzerland

BACKGROUND & AIMS: **Eosinophilic esophagitis (EoE) is assessed endoscopically (endoscopic activity), based on grades of edema, rings, exudates, furrows, and strictures (EREFS). We examined variations in endoscopic assessments of severity, developed and validated 3 EREFS-based scoring systems, and assessed responsiveness of these systems using data from a randomized placebo-controlled trial of patients with EoE.**

METHODS: **For the development set, 5 gastroenterologists reviewed EREFS findings from 266 adults with EoE and provided endoscopist global assessment scores (EndoGA, scale of 0 to 10); variation (ΔEndoGA) was assessed using linear regression. We evaluated simple scores (features given arbitrary values from 0 to 3) and developed 2 scoring systems (adjusted score range, 0–100). We then fitted our linear regression model with mean EndoGA to data from 146 adults recruited in centers in Switzerland and the United States between April 2011 and December**

[a]Authors share co-first authorship.

Abbreviations used in this paper: CI, confidence interval; EEsAI, eosinophilic esophagitis activity index; EGD, esophagogastroduodenoscopy; EndoGA, endoscopist global assessment; EoE, eosinophilic esophagitis; EREFS, edema, rings, exudates, furrows, and strictures; IQR, interquartile range.

Most current article

Dr. Falk Pharma v. Ellodi Pharmaceuticals
IPR2025-00054
Dr. Falk Pharma Ex. 1073

© 2019 by the AGA Institute
1542-3565/$36.00
https://doi.org/10.1016/j.cgh.2018.11.032

2012. For the validation set, we collected data from 120 separate adults (recruited in centers in Switzerland and the United States between May 2013 and July 2014), assessing regression coefficient-based scores using Bland-Altman method. We assessed the responsiveness of our scoring systems using data from a randomized trial of patients with EoE given fluticasone (n=16) or placebo (n=8).

RESULTS: The distribution of EndoGA values differed among endoscopists (mean ΔEndoGA, 2.6±1.8; range 0–6.6). We developed 2 regression-based scoring systems to assess overall and proximal and distal esophageal findings; variation in endoscopic features accounted for more than 90% of the mean EndoGA variation. In the validation group, differences between mean EndoGA and regression-based scores were small (ranging from −4.70 to 2.03), indicating good agreement. In analyses of data from the randomized trial, the baseline to end of study change in patients given fluticasone was a reduction of 24.3 in simple score (reduction of 4.6 in patients given placebo, $P=.052$); a reduction of 23.5 in regression-based overall score (reduction of 6.56 in patients given placebo, $P=.12$), and a reduction of 23.8 (reduction of 8.44 in patients given placebo, $P=.11$).

CONCLUSION: Assessments of endoscopic activity in patients with EoE vary among endoscopists. In an analysis of data from a randomized controlled trial, we found that newly developed scoring systems are no better than simple scoring system in detecting changes in endoscopic activity. These results support the use of a simple scoring system in evaluation of endoscopic activity in patients with EoE. clinicaltrials.gov no: NCT00939263 and NCT 01386112

Keywords: Index; Esophagus; Variability in Endoscopic Assessment; Instrument.

Eosinophilic esophagitis (EoE) is defined as "a chronic, immune/antigen-mediated, esophageal disease characterized clinically by symptoms related to esophageal dysfunction and histologically by eosinophil-predominant inflammation".[1–3]

Because dysphagia is the most frequent complaint of adult EoE patients, esophagogastroduodenoscopy (EGD) represents an important diagnostic procedure.[1–3] Although endoscopic abnormalities are not pathognomonic of EoE, these are frequently assessed to aid in clinical decision-making and interpretation of results of clinical trials and observational studies. In 2013, Hirano et al[4] described a classification and grading system of the following EoE-associated endoscopic alterations: edema, rings, exudates, furrows, and stricture(s) (EREFS). Although no score to assess the overall endoscopic activity was developed, attempts have been made to use the EREFS system as a basis for a simple score, which is calculated by summing together the arbitrary values from 0 to 3 on the basis of presence and severity of various EoE-associated features.[5,6] Dellon et al[7] reported that giving more weight to inflammatory features such as exudates and edema renders the EREFS-based score more responsive to anti-inflammatory treatment. These studies revealed differences in the way gastroenterologists synthesize information about the severity of individual features when assessing overall endoscopic severity and emphasize the importance of methodologic considerations in developing an endoscopic score.

Therefore, we used data from the adult EoE activity index (EEsAI) study population to explore the variation in gastroenterologists' assessment of endoscopic severity, used this information to develop and validate 3 EREFS-based scores, and assessed responsiveness of these scores by using the data from randomized placebo-controlled clinical trial of fluticasone.[8]

Methods

Study Population

EEsAI study (clinicaltrials.gov ID: NCT00939263) was approved by local institutional review boards. The patient inclusion and exclusion criteria were previously published.[9]

Assessment of Endoscopic Findings

Distal and proximal esophageal sections were defined as those spanning 5 cm above the gastroesophageal junction and the top one-half, respectively. Patients underwent EGD, during which endoscopic findings were assessed by using EREFS system with modifications for stricture definitions (Supplementary Table 1).[4] The endoscopists performing EGD also provided the overall severity grading (absent, mild, moderate, severe).

Five endoscopists (A.S., A.M.S., E.S.D., I.H., Y.R., who performed ≥200 EGDs in EoE patients) were provided a datasheet containing the grading of inflammatory (white exudates, furrows, edema) and fibrotic (rings, strictures, crêpe-paper) features. The experts provided an Endoscopist Global Assessment (EndoGA) ranging from 0 to

10 (0, inactive EoE; 10, most active EoE) as an overall impression of endoscopic severity.

Data Handling and Statistical Analysis

Data were double-entered into EpiData (version 3.1) and imported into Stata (version 13.1; StataCorp, College Station, TX) or R Project for analyses. Results are presented as percentages for categorical variables or median (interquartile range [IQR]) for continuous variables. The most severe category for a given feature found in either esophageal section represented overall severity. If data on feature's severity in one esophageal section were missing, then that feature's severity in another section represented overall severity.

Relationship Between Severity Grading and Endoscopist Global Assessment

We analyzed the relationship between EndoGA (outcome) and overall severity grading (fixed effect) by using linear mixed-effects regression. We examined box plots of mean EndoGA for EGD performed in one's own center ($EndoGA^{own}$), in another center ($EndoGA^{others}$), and mean difference between $EndoGA^{own}$ and $EndoGA^{others}$ versus overall severity grading.

Variation in Expert Assessment of Endoscopic Severity

We examined the variation in the way endoscopists assessed overall endoscopic severity by using several approaches. (1) The probability distributions of all EndoGA values provided by each expert were visualized by using violin plots and box plots. (2) To examine the per-patient variation, we calculated the difference between highest and lowest EndoGA values (ΔEndoGA) from among 5 values assigned to findings of single EGD. A large ΔEndoGA value reflects disagreement in assessing that patient's endoscopic severity. To examine which features contributed most to per-patient variation, ΔEndoGA was regressed on each feature, and the coefficient of determination (R^2) was calculated. Features with a higher R^2 are more strongly associated with ΔEndoGA variance. A multivariable regression model with ΔEndoGA as outcome and all the features as predictors was used to examine each feature's contribution to variation in ΔEndoGA adjusted for all other features. (3) To assess the per-feature variation, we fitted a multivariable regression model to each expert's EndoGA values as outcome and the features as independent variables.

Development and Validation of Scores

We developed 3 EREFS-based scores (Table 1). For each patient, the mean of 5 EndoGA values was used as outcome.

What You Need to Know

Background

In eosinophilic esophagitis, endoscopic activity is graded by examining the presence and severity of Edema, Rings, Exudates, Furrows, and Strictures (EREFS). As there is paucity of data, we examined variation in the way experts assessed endoscopic severity, developed and validated three EREFS-based scores, and evaluated the scores' responsiveness in clinical trial of fluticasone.

Findings

The endoscopic severity impression differs among expert endoscopists. Exudates accounted for most variation in severity assessment. The responsiveness of new scores considering expert opinion was no better than that of simple score (features given arbitrary values from 0-3), when clinical trial data were analyzed.

Implications for patient care

The simple EREFS score should be used in short-term clinical trials of anti-inflammatory therapies. The new score should be examined in long-term, observational studies of patients with broader endoscopic severity spectrum.

Using the evaluation group data, the simple EREFS score ranging from 0 to 8 was calculated (Supplementary Table 1). The scores were used as predictors for mean EndoGA in a linear regression, and the goodness of fit was assessed by using R^2.

Using the evaluation group data, we developed 2 scores, the weighted EREFS score and the weighted EREFS-proximal/distal score, by fitting the predictors to the mean EndoGA scores using linear regression models. We assessed goodness of fit by using R^2. The models' coefficients were used as values for the scores, which were transformed to range 0–100 (coefficients×100/∑all coefficients, rounded to 0.5). For weighted EREFS-proximal/distal score, most EREFS features were graded as described by Hirano et al.[4] The presence of crêpe-paper and the stricture severity were also examined. We also created the binary variables to indicate that most severe form of a feature is present in both esophageal sections. We removed nonsignificant predictors from the model by backward elimination process. We also developed weighted scores using $EndoGA^{own}$ as outcome.

To validate weighted scores (based on EndoGA or $EndoGA^{own}$), we used validation group data. The agreement between the EndoGA/$EndoGA^{own}$ and the calculated score values was examined by using the following plots: (1) Bland-Altman: (10×EndoGA-Score) versus (10×EndoGA+Score)/2 (the closer the horizontal line is to 0, the better is agreement between the measures); and (2) calibration: EndoGA/$EndoGA^{own}$ versus score values

Table 1. Summary of the Developed Scores

	Simple EREFS score	Weighted EREFS score	Weighted EREFS-proximal/distal score
Takes into account expert considerations	No	Yes	Yes
Statistical methods for score derivation	Not applicable	Linear regression with mean of 5 EndoGA values as outcome	Linear regression with mean of 5 EndoGA values as outcome
Worst endoscopic presentation represents overall endoscopic severity for that feature	Yes	Yes	No, severity of endoscopic features is assessed in both proximal and distal esophagus by introducing extra variables denoting presence of endoscopic features in both parts of the esophagus
Validation	Not applicable	Validated in second independent group of patients	Validated in second independent group of patients
Responsiveness	Yes, evaluated using data from short-term randomized placebo-controlled clinical study of fluticasone	Yes, evaluated using data from short-term randomized placebo-controlled clinical study of fluticasone	Yes, evaluated using data from short-term randomized placebo-controlled clinical study of fluticasone
Components of the score	Rings None Mild Moderate Severe Exudates None Mild Severe Furrows Absent Present Edema Absent Present Strictures Absent Present Crêpe-paper Absent Present	Rings None Mild Moderate Severe Exudates None Mild Severe Furrows Absent Present Edema Absent Present Strictures Absent Present Crêpe-paper Absent Present	Rings None Mild in prox and/or dist Moderate in prox and/or dist Severe in prox and/or dist Exudates None Mild in prox and/or dist Severe in prox and/or dist Edema Absent Present in prox and/or dist Crêpe paper Absent Present in prox and/or dist Strictures Absent Low-grade in prox and/or dist Intermediate/high in prox and/or dist Severe rings Absent in prox and dist Present in prox and dist Severe exudates Absent in prox and dist Present in prox and dist Furrows Absent in prox and dist Present in prox and dist Edema Absent in prox and dist Present in prox and dist

dist, distal part of esophagus; EndoGA, endoscopist global assessment; EREFS, edema, rings, exudates, furrows, and strictures; prox, proximal part of esophagus.

(goodness of fit evaluated by using R^2; for a perfect score the line fitted between EndoGA/EndoGA^{own} and score values has a slope of 1). For each score, we also fit a linear regression model with score value as predictor and EndoGA^{own} as outcome. R^2 was calculated and meta-analyzed by using a random-effects meta-analysis model (R's "metafor" package). We estimated the R^2 variance by using the CI.Rsq function (R's "psychometric" package) and calculated the standard error/confidence intervals (CIs) of R^2 using a large sample approximation.[9]

Fluticasone Clinical Trial and Scores' Responsiveness

A randomized, double-blind, placebo-controlled trial (Phase 1b/2a) of fluticasone (APT-1011) examined the tolerability and safety of 2 APT-1011 dosing regimens compared with placebo in adolescent/adult EoE patients (clinicaltrials.gov ID: NCT01386112).[8] Subjects were randomized 1:1:1 to receive APT-1011 1.5 mg twice a day (n = 8), APT-1011 3.0 mg every day (n = 8), or placebo (n = 8). During EGDs at baseline and end of treatment (week 8), EREFS were assessed. At end of treatment, the median esophageal eosinophil counts/mm^2 were significantly decreased from baseline in biopsies of patients given APT-1011 but not placebo (379 [IQR, 289–563] to 0 [IQR, 0–60] in APT-1011 1.5 mg twice a day; 378 [IQR, 224–458] to 23 [IQR, 0–109] in APT-1011 3.0 mg every day; and 459 [IQR, 286–609] to 323 [IQR, 200–523] in placebo). We examined scores' change (responsiveness) from baseline to end of treatment in drug-treated groups pulled together relative to placebo by using analysis of covariance models with time, treatment group, and an interaction term as fixed effects.

All authors had access to the study data and reviewed and approved the final manuscript.

Results

The characteristics of 266 prospectively included adult EoE patients are shown in Table 2. The endoscopic characteristics of patients are shown in Supplementary Table 2.

Relationship Between Endoscopist Global Assessment and Overall Severity Grading

We explored the relationship between overall severity grading and EndoGA (Supplementary Figure 1). We observed a 2-point increase in predicted EndoGA for 1-level increase in a severity grading. This increase was independent of whether a physician performed EGD or ranked the EREFS findings provided by another physician. This is further confirmed by our finding that the mean difference between EndoGA^{own} and EndoGA^{others} values is centered around 0, irrespective of the severity grading.

Variation in Endoscopic Activity Assessment

When examining per-expert variation in EndoGA value distribution in the evaluation group (Figure 1*A*), we found that experts 3, 4, and 5 have a higher number of EndoGA values at the lower spectrum of severity. The mean EndoGA values (± standard deviation and range) provided by endoscopists were 4.35 ± 2.91 (0–10), 3.55 ± 2.18 (0–9), 2.27 ± 1.96 (0–8.54), 3.51 ± 2.46 (0–10), and 3.44 ± 2.50 (0–10).

The median ΔEndoGA for each patient's EGD was 2.29 (IQR, 1.02–3.61; range, 0–6.56). To evaluate which endoscopic features contributed most to ΔEndoGA variation, we examined the univariable linear regression R^2 values, which were as follows: exudates, $R^2 = 0.411$; rings, $R^2 = 0.128$; edema, $R^2 = 0.347$; furrows, $R^2 = 0.299$; strictures, $R^2 = 0.002$; stricture diameter, $R^2 = 0.026$; and crêpe-paper, $R^2 = 0.043$. Thus, exudates, edema, and furrows explained a relevant part of ΔEndoGA variation. Using multivariable analyses, we found that except for crêpe-paper ($P = .634$), all the other features explained a part of ΔEndoGA variation (exudates, $P < .001$; rings, $P = .003$; edema, $P < .001$; furrows, $P = .048$; and stricture diameter, $P = .03$).

Last, we analyzed the variation in the way experts synthesize information about the presence/severity of endoscopic features by regressing each expert's EndoGA on the features (Supplementary Table 3, Figure 1*B*). The experts "attributed" similar weights to the presence of furrows, edema, and crêpe-paper. For features with more than 2 severity levels, we observed that the more severe the feature, the greater is the variation in the expert-attributed weights.

Development and Validation of Endoscopic Scores

Simple edema, rings, exudates, furrows, and strictures scores. The median simple EREFS score in evaluation group was 3 (IQR, 2–5; range, 0–8). Using linear regression modelling in evaluation group, we found that a simple EREFS score explained 90.0% variation in the mean EndoGA ($R^2 = 0.900$; coefficient = 1.100; 95% CI, 1.040–1.161; intercept = −0.44). The crêpe-paper addition did not improve this performance. The correlations between the mean EndoGA values and simple EREFS-based scores are shown in Supplementary Figure 2.

Weighted edema, rings, exudates, furrows, and strictures scores. We then developed an EREFS-based score estimating the weights that expert endoscopists attributed to various features (Table 3). The variation in the severity of major endoscopic features explained 91.6% variation in EndoGA. The model's coefficients were used as the values for the score (transformed to range 0–100). In the validation group, the weighted EREFS score explained 89.2% of the mean EndoGA variation (Figure 2*A*). In Bland-Altman plot, the mean difference of −4.70 between weighted EREFS score and the mean EndoGA was observed (95% CI, 10.46 to −19.86) (Figure 2*B*). The crêpe-paper addition had a small impact on weighted score (Figure 2*C* and *D*). If validation was carried out by using EndoGA^{own} as an outcome, the weighted EREFS score explained 80.0% (meta-analyzed R^2; Supplementary Figure 3*B*) and 77.1% (Supplementary Figure 4*A*) of the mean EndoGA^{own} variation. In Bland-Altman analyses, the mean difference of −3.54 between weighted EREFS score and the mean EndoGA^{own} was observed (95% CI, 20.74 to −27.81) (Supplementary Figure 4*B*).

Table 2. Patient Characteristics

Characteristic	Evaluation group		Validation group	
	Frequency	%	Frequency	%
No. of patients	146	(100.0)	120	(100.0)
Male	104	(71.2)	73	(60.8)
Age at inclusion, *y* (median, interquartile range, range)	37.7	(29–46, 18–71)	40.5	(31–49, 19–80)
Ethnicity				
White	142	(97.3)	114	(95.0)
Non-white	4	(2.7)	6	(5.0)
Education				
Compulsory schooling	2	(1.4)	1	(0.8)
Vocational training	36	(24.7)	33	(27.5)
Upper secondary education	63	(43.2)	54	(45.0)
University education	45	(30.8)	32	(26.7)
Eosinophilic esophagitis symptoms onset				
1–11 mo ago	8	(5.5)	2	(1.7)
1–5 y ago	61	(41.8)	38	(31.7)
>5 y ago	77	(52.7)	80	(66.6)
Clinical activity (EEsAI PRO score, range 0–100)	27	(12–46, 0–94)	27	(6–42, 0–94)
Peak esophageal eosinophil count per mm^2 (median, interquartile range, range)	95	(26–226, 0–744)	87	(11–305, 0–1558)
Allergic diseases/allergies				
Asthma	53	(36.3)	42	(35.0)
Rhinoconjunctivitis	87	(59.6)	72	(60.0)
Eczema	18	(12.3)	34	(28.3)
Food allergy	43	(29.5)	60	(50.0)
Gastroesophageal reflux disease	45	(30.8)	18	(15.0)
Diagnosis established				
Clinically	26	(57.8)	3	(16.7)
Endoscopically	11	(24.4)	6	(33.3)
Based on pH-metric studies	1	(2.2)	2	(11.1)
Clinically and endoscopically	8	(17.8)	5	(27.8)
Concomitant medications				
Proton pump inhibitors	80	(54.8)	39	(32.5)
Histamine antagonists (H_2-receptor)	5	(3.4)	1	(0.8)
Histamine antagonists (H_1-receptor)	24	(16.4)	18	(15.0)
Inhaled corticosteroids for asthma	4	(2.7)	4	(3.3)
β2-adrenergic agonists for asthma	21	(14.4)	2	(1.7)
Leukotriene receptor antagonists for asthma	4	(2.7)	1	(0.8)
EoE-specific treatments in the last 12 mo	86	(58.9)	103	(85.8)
Hypoallergenic diets	17	(11.6)	19	(15.8)
Swallowed topical corticosteroids	62	(42.5)	78	(65.0)
Esophageal dilation	27	(18.5)	26	(21.7)

EEsAI, eosinophilic esophagitis activity index; PRO, patient-reported outcomes.

Weighted edema, rings, exudates, furrows, and strictures–proximal/distal score. We developed a weighted EREFS score that considered stricture severity and presence of crêpe-paper. The variation in the severity of different features explained 95.9% variation in EndoGA ($R^2 = 0.959$). Using backwards elimination, crêpe-paper and furrowing were removed, with only a minor change in R^2 (Table 4). The final model's coefficients were used as the values for the score (transformed to range 0–100). In the validation group, the weighted EREFS-proximal/distal score estimated 94.8% of the mean EndoGA (Figure 2*E*). In Bland-Altman plot, mean difference of 2.03 between weighted EREFS-proximal/distal score and the mean EndoGA was observed (95% CI, 12.12 to −8.05) (Figure 2*F*). If validation was carried out using $EndoGA^{own}$ as an outcome, the weighted EREFS-proximal/distal score explained 85% of the mean $EndoGA^{own}$ variation (meta-analyzed R; Supplementary Figure 3*C*, Supplementary Figure 4*C*). In Bland-Altman analyses, the mean difference of 3.46 between weighted EREFS-proximal/distal score and the mean $EndoGA^{own}$ was observed (95% CI, 24.12 to −17.21) (Supplementary Figure 4*D*).

The data on development and validation of the weighted scores based on $EndoGA^{own}$ as outcome are shown in Supplementary Tables 4 and 5, Supplementary Figure 5.

Scores' responsiveness. The rate of baseline to end-of-treatment change in fluticasone-treated patients (n = 16) (relative to placebo [n = 8]) was −24.3 (−4.6, $P = .052$), −23.5 (−6.56, $P = .12$), and −23.8 (−8.44, $P = .11$) for simple EREFS score, weighted EREFS score, and weighted EREFS-proximal/distal scores, respectively (Supplementary Table 6). We separately examined the

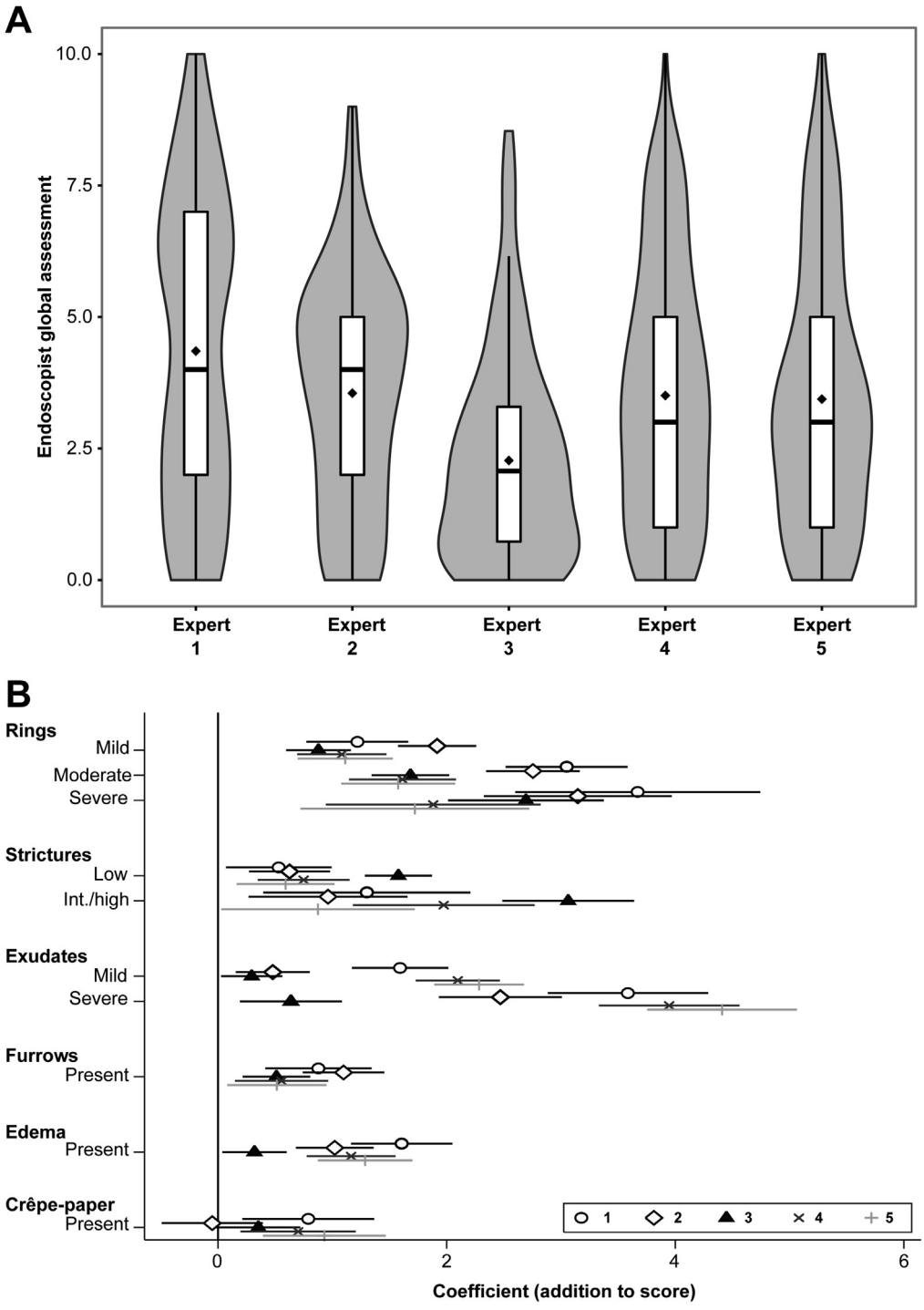


Figure 1. (*A*) Distribution of global assessment values provided by endoscopists. *Vertical lines* indicate interquartile range; *crossing horizontal line* is at the median. *Rhombi* indicate the mean. (*B*) Variation in the weights that 5 endoscopists attribute to different endoscopic features.

effect treatment had on inflammatory and fibrotic features. The slope analysis detected significant treatment benefit over placebo on inflammatory features, when using simple but not weighted scores.

Discussion

Because consensus on grading endoscopic severity in currently published studies is lacking, we evaluated 3 EREFS-based scores for endoscopic activity assessment in adults with EoE. We found that (1) overall impression of EoE endoscopic severity differs among endoscopists; (2) the simple score explained the overall variation of endoscopic severity to a similar extent compared with the weighted scores; and (3) when evaluating scores' responsiveness in a clinical trial, the weighted scores were statistically no better than the simple EREFS score in differentiating fluticasone-treated from placebo-treated patients. Hence, the simple score should be used in short-term clinical trials of anti-inflammatory therapies. Whether weighted EREFS-proximal/distal

Table 3. Multivariable Linear Regression Model for Derivation of Weighted EREFS Scores (With and Without Crêpe-paper): Final Scores

	Coefficient[a]	95% CI	*P* value	Score (total set to 100)	Coefficient[a]	95% CI	*P* value	Score (total set to 100)
Rings								
None	0.000	Ref		0	0.000	Ref		0
Mild	1.239	0.941–1.538	<.001	13	1.223	0.937–1.510	<.001	12.5
Moderate	2.169	1.807–2.531	<.001	23	2.132	1.784–2.480	<.001	22
Severe	3.332	2.773–3.890	<.001	35.5	3.216	2.677–3.756	<.001	33.5
Exudates								
None	0.000	Ref		0	0.000	Ref		0
Mild	1.341	1.058–1.624	<.001	14	1.373	1.101–1.645	<.001	14
Severe	3.155	2.685–3.625	<.001	33.5	3.082	2.629–3.534	<.001	32
Furrows								
Absent	0.000	Ref		0	0.000	Ref		0
Present	0.586	0.277–0.896	<.001	6	0.608	0.311–0.906	<.001	6.5
Edema								
Absent	0.000	Ref		0	0.000	Ref		0
Present	1.232	0.939–1.524	<.001	13	1.128	0.841–1.414	<.001	11.5
Strictures								
Absent	0.000	Ref		0	0.000	Ref		0
Present	1.117	0.822–1.411	<.001	12	0.965	0.671–1.260	<.001	10
Crêpe-paper	NA	NA	NA	NA				
Absent					0.000	Ref		0
Present					0.652	0.292–1.013	<.001	6.5
Constant[b]	0.044	–0.201 to 0.289	.722		0.078	–0.158 to 0.314	.514	
R^{2}[c]	0.916				0.923			
Sum of the coefficients/score	9.422			100	9.651			100

CI, confidence interval; EndoGA, endoscopist global assessment; EREFS, edema, rings, exudates, furrows, and strictures; Ref, reference.
[a]Coefficient represents the EndoGA value change for each endoscopic feature. For example, mean EndoGA increased by 1.239, if mild rings were found. In these analyses, the adjusted regression coefficient for rings represents the amount of mean EndoGA variation that is owing to the rings alone, after the presence of all other features was considered. If mild rings and edema were detected, then mean EndoGA increased by 2.471 (1.239 for mild rings and 1.232 for edema).
[b]Constant represents the predicted EndoGA value, when all values of independent variables are set to reference category.
[c]R^2 is a measure of the extent to which the regression model describes the data. The closer R^2 is to 1, the more precise the regression model is.

score should be used in long-term trials and studies remains to be elucidated. As for clinical practice, we advocate for the use of simple score, because the convenience of using it outweighs the modest gains in precision provided by newly developed scores.

The study's findings have important implications for endoscopic activity assessment. Exudates contributed most to the per-endoscopist variation in activity assessment. Compared with simple score, new scores put more weight on exudates and less on the furrows, which is consistent with symptom severity in the same population.[10] We observed no benefit in scoring crêpe-paper, whereas inclusion of stricture diameter is of importance, because persons with intermediate-/high-grade strictures experience severe symptoms and diminished EoE-specific quality of life.[10,11] Values attributed to furrows and edema should not be doubled, as previously suggested.[7] Inclusion of variables that account for the endoscopic severity in both esophageal sections resulted in statistically best score. Whether assessment of endoscopic findings in both esophageal sections is of benefit remains to be determined. The weighted scores derived on the basis of endoscopic activity assessment of the physician performing EGD are statistically more heterogeneous and likely to provide less consistent results, when compared with the scores derived on the basis of expert group judgment.

The optimal approach to assessing furrows severity using EREFS system is a subject of much research. Hirano et al[4] reported that interobserver agreement for distinguishing between mild and severe furrows was moderate; hence, these categories were collapsed for the final EREFS system. Nevertheless, both types of furrows severity grading have been previously reported. Dellon et al[6,7] graded furrows severity as absent, mild, and severe in both an observational study and budesonide clinical trial. In contrast, van Rhijn et al[5] graded furrows as absent/present. These studies also evaluated EREFS-based scores' responsiveness to treatment.[5–7] Because many EoE therapies are anti-inflammatory, researchers increased the weight of inflammatory features. For example, Dellon et al[7] demonstrated that doubling the points attributed to exudates and edema improved the score responsiveness.

Therefore, we developed weighted EREFS scores that incorporate expert considerations in a population larger than used previously (266 compared with 93, 67, and 69 patients, respectively).[5–7] Compared with the

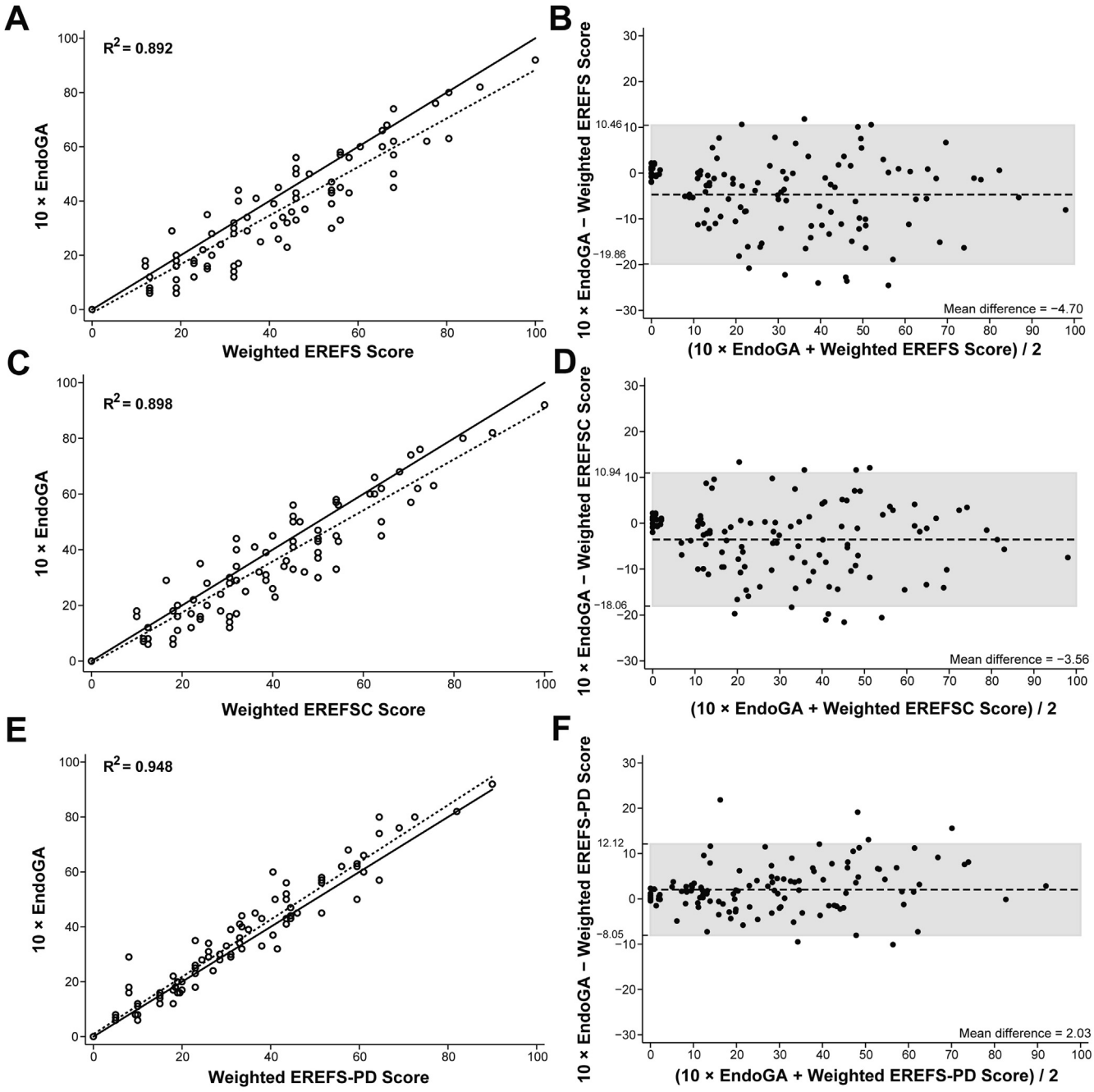


Figure 2. Calibration plots for (*A*) weighted EREFS, (*C*) weighted EREFS with crêpe-paper (EREFSC), and (*E*) weighted EREFS-proximal/distal (EREFS-PD) scores in the validation group. *Solid line* has a slope of 1 and represents an ideal relationship between a perfect score and EndoGA values. *Dashed line* is the regression line fit to the data. Bland-Altman plots for the agreement between (*B*) weighted EREFS, (*D*) weighted EREFSC, and (*F*) weighted EREFS-PD scores and 10×EndoGA in the validation group. *Gray box* indicates the 95% confidence intervals limits of agreement. EndoGA, endoscopist global assessment; EREFS, edema, rings, exudates, furrows, and strictures.

simple score, experts weighted severe exudates more than in previous studies but furrows and edema similarly. Our data are similar to those on symptom severity, because patients with severe exudates experience worse EoE-specific quality of life and severe symptoms.[10,11] In contrast, quality of life and symptoms of patients in whom furrows and edema are present do not differ from those of patients lacking these features.[9,10] We found that weighted EREFS-proximal/distal score correlates best with EndoGA and has the smallest mean difference and limits of agreement, but it was no better at detecting fluticasone treatment benefit than a simple score.

Endoscopists provided EndoGA by examining a datasheet containing the summary of endoscopic findings to prevent the ambiguities in EREFS ranking, which would preclude us from developing a score (such as equation with 12 unknowns). However, experts' EndoGA

Table 4. Multivariable Linear Regression Model for Derivation of the Weighted EREFS-Proximal/Distal Score: Final Score

	Model 1			Model 2			Model 3			Score
	Coeff[a]	95% CI	*P* value	Coeff[a]	95% CI	*P* value	Coeff[a]	95% CI	*P* value	(total set to 100)
Rings										
None	0.000	Ref	<.001	0.000	Ref	<.001	0.000	Ref	<.001	0
Mild in prox and/or dist	1.018	0.796–1.241		1.018	0.796–1.240		1.022	0.801–1.242		10
Moderate in prox and/or dist	1.871	1.605–2.136		1.871	1.607–2.136		1.875	1.612–2.138		18
Severe in prox and/or dist	1.703	1.043–2.362		1.678	1.034–2.322		1.677	1.034–2.319		18
Exudates										
None	0.000	Ref	<.001	0.000	Ref	<.001	0.000	Ref	<.001	0
Mild in prox and/or dist	1.525	1.316–1.734		1.524	1.316–1.732		1.537	1.337–1.738		15
Severe in prox and/or dist	2.021	1.402–2.639		2.042	1.436–2.648		2.055	1.453–2.657		19.5
Furrows							—	—	—	—
Absent	0.000	Ref	.653	0.000	Ref	.633				
Present in prox and/or dist	0.082	–0.278 to 0.441		0.087	–0.271 to 0.444					
Edema										
Absent	0.000	Ref	.048	0.000	Ref	.050	0.000	Ref	.009	0
Present in prox and/or dist	0.475	0.005–0.944		0.464	0.000–0.929		0.526	0.136–0.915		5
Crêpe-paper										
Absent	0.000	Ref	045	0.000	Ref	<.001	0.000	Ref	<.001	0
Present in prox and/or dist	0.471	0.011–0.931		0.540	0.254–0.826		0.534	0.250–0.818		5
Strictures										
Absent	0.000	Ref	<.001	0.000	Ref	<.001	0.000	Ref	<.001	0
Low-grade in prox and/or dist	0.836	0.612–1.060		0.835	0.612–1.058		0.836	0.613–1.058		8
Intermediate/high in prox and/or dist	1.875	1.426–2.324		1.872	1.425–2.319		1.866	1.421–2.311		18
Severe rings										
Absent in prox and dist	0.000	Ref	.017	0.000	Ref	.013	0.000	Ref	.011	0
Present in prox and dist	0.776	0.141–1.410		0.795	0.170–1.419		0.808	0.188–1.429		8
Severe exudates										
Absent in prox and dist	0.000	Ref	<.001	0.000	Ref	<.001	0.000	Ref	<.001	0
Present in prox and dist	1.377	0.650–2.104		1.342	0.641–2.043		1.333	0.635–2.031		13
Furrows										
Absent in prox and dist	0.000	Ref	<.001	0.000	Ref	<.001	0.000	Ref	<.001	0
Present in prox and dist	0.893	0.539–1.248		0.891	0.538–1.244		0.958	0.735–1.181		9
Edema										
Absent in prox and dist	0.000	Ref	.039	0.000	Ref	.031	0.000	Ref	.027	0
Present in prox and dist	0.508	0.025–0.990		0.522	0.047–0.997		0.465	0.055–0.875		4.5
Crêpe-paper				—	—	—	—	—	—	—
Absent in prox and dist	0.000	Ref	.707							
Present in prox and dist	0.101	–0.429 to 0.630								
Constant	0.193	0.131–0.374	.036	0.192	0.122–0.371	.037	0.201	0.027–0.376	.024	—
R^2	0.959			0.959			0.958			
Sum							10.419			100

NOTE. Model 1 includes all predictors, whereas models 2 and 3 are fitted to the data after exclusion of the least significant features.
Coeff, coefficient; CI, confidence interval; dist, distal; EndoGA, endoscopist global assessment; EREFS, edema, rings, exudates, furrows, and strictures; prox, proximal; Ref, reference.
[a]Coefficient represents the EndoGA value change for each endoscopic feature. For example, mean EndoGA increased by 1.703 if severe rings in proximal and/or distal esophagus were found (Model 1). In these analyses, the adjusted regression coefficient for rings represents the amount of mean EndoGA variation that is owing to rings alone, after the presence of all other features was considered.

was strongly associated with the severity grading obtained directly after the EGD, and the increase in EndoGA associated with increase in the severity grading was independent of whether physician performed EGD or ranked the EREFS findings provided by another physician.

Our study has strengths and limitations. The data provided to the experts included all EREFS features, and these were used as predictors for score development; therefore, as expected, large part of the variance is explained by these variables. In our study, the stricture diameter was estimated relative to endoscope diameter. Future studies should assess the stricture diameter by using balloons, bougies, or the functional luminal imaging probe. The scores' responsiveness was evaluated by using the data from the trial, into which patients with relatively mild disease activity were included. We wonder whether if weighted scores were to perform differently, patients with more severe disease were included. Nevertheless, the study's limitations are

countered by strengths including a large sample size, inclusion of independent group for score validation, attempt to seek consensus among experts, and scores' responsiveness assessment by using clinical trial data.

In summary, we examined various scores for endoscopic activity assessment in adult EoE patients. The new scores were no better in detecting the fluticasone treatment benefit over placebo compared with a simple EREFS score. These results suggest that the simple EREFS score should be used in short-term clinical trials of anti-inflammatory therapies.

Supplementary Material

Note: To access the supplementary material accompanying this article, visit the online version of *Clinical Gastroenterology and Hepatology* at www.cghjournal.org, and at https://doi.org/10.1016/j.cgh.2018.11.032.

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. Lucendo AJ, Molina-Infante J, Arias Á, et al. Guidelines on eosinophilic esophagitis: evidence-based statements and recommendations for diagnosis and management in children and adults. United European Gastroenterol J 2017;5:335–358.
3. Furuta GT, Katzka DA. Eosinophilic esophagitis. N Engl J Med 2015;373:1640–1648.
4. Hirano I, Moy N, Heckman MG, et al. Endoscopic assessment of the oesophageal features of eosinophilic esophagitis: validation of a novel classification and grading system. Gut 2013;62:489–495.
5. van Rhijn BD, Verheij J, Smout AJ, et al. The Endoscopic Reference Score shows modest accuracy to predict histologic remission in adult patients with eosinophilic esophagitis. Neurogastroenterol Motil 2016;28:1714–1722.
6. Dellon ES, Katzka DA, Collins MH, et al. Budesonide oral suspension improves symptomatic, endoscopic, and histologic parameters compared with placebo in patients with eosinophilic esophagitis. Gastroenterology 2017;152:776–786.
7. Dellon ES, Cotton CC, Gebhart JH, et al. Accuracy of the eosinophilic esophagitis endoscopic reference score in diagnosis and determining response to treatment. Clin Gastroenterol Hepatol 2016;14:31–39.
8. Hirano I, Schoepfer AM, Comer GM, et al. A randomized, double-blind, placebo-controlled trial of a fluticasone propionate orally disintegrating tablet in adult and adolescent patients with eosinophilic esophagitis: a Phase 1/2A safety and tolerability study. Gastroenterology 2017;152:S195.
9. Cohen J, Cohen P, West SG, et al. Applied multiple regression/correlation analysis for the behavioral sciences. Mahwah, NJ: Lawrence Erlbaum Associates, 2003.
10. Schoepfer AM, Straumann A, Panczak R, et al. Development and validation of a symptom-based activity index for adults with eosinophilic esophagitis. Gastroenterology 2014;147:1255–1266.
11. Safroneeva E, Coslovsky M, Kuehni CE, et al. Eosinophilic oesophagitis: relationship of quality of life with clinical, endoscopic and histological activity. Aliment Pharmacol Ther 2015;42:1000–1010.

Reprint requests

Address requests for reprints to: Alain Schoepfer, MD, Division of Gastroenterology and Hepatology, Centre Hospitalier Universitaire Vaudois/CHUV and University of Lausanne, Rue de Bugnon 44, 07/2425, 1011 Lausanne, Switzerland. e-mail: alain.schoepfer@chuv.ch; fax: + 41-21-314-47-18.

Acknowledgments

Members of the International EEsAI study group participating in data collection (in alphabetical order): Sami R. Achem (Mayo Clinic, Jacksonville, FL), Amindra S. Arora (Mayo Clinic, Rochester, MN), Oral Alpan (O&O Alpan, LLC, Section on Immunopathogenesis, Fairfax, VA), David Armstrong (McMaster University, Hamilton, Canada), Stephen E. Attwood (North Tyneside Hospital, North Shields, UK), Joseph H. Butterfield (Mayo Clinic, Rochester, MN), Michael D. Crowell (Mayo Clinic, Scottsdale, AZ), Kenneth R. DeVault (Mayo Clinic, Jacksonville, FL), Eric Drouin (CHU Sainte-Justine, Montreal, Canada), Benjamin Enav (Pediatric Gastroenterology of Northern Virginia, Fairfax, VA), Felicity T. Enders (Mayo Clinic, Rochester, MN), David E. Fleischer (Mayo Clinic, Scottsdale, AZ), Amy Foxx-Orenstein (Mayo Clinic, Scottsdale, AZ), Dawn L. Francis (Mayo Clinic, Jacksonville, FL), Gordon H. Guyatt (McMaster University, Hamilton, Canada), Lucinda A. Harris (Mayo Clinic, Scottsdale, AZ), Amir F. Kagalwalla (Northwestern University Feinberg School of Medicine, Chicago, IL), Hirohito Kita (Mayo Clinic, Rochester, MN), Murli Krishna (Mayo Clinic, Jacksonville, FL), James J. Lee (Mayo Clinic, Scottsdale, AZ), John C. Lewis (Mayo Clinic, Scottsdale, AZ), Kaiser Lim (Mayo Clinic, Rochester, MN), G. Richard Locke III (Mayo Clinic, Rochester, MN), Joseph A. Murray (Mayo Clinic, Rochester, MN), Cuong C. Nguyen (Mayo Clinic, Scottsdale, AZ), Diana M. Orbelo (Mayo Clinic, Rochester, MN), Shabana F. Pasha (Mayo Clinic, Scottsdale, AZ), Francisco C. Ramirez (Mayo Clinic, Scottsdale, AZ), Javed Sheikh (Kaiser Permanente Los Angeles Medical Center, Los Angeles, CA), Sarah B. Umar (Mayo Clinic, Scottsdale, AZ), Catherine R. Weiler (Mayo Clinic, Rochester, MN), John M. Wo (Indiana University, Indianapolis, IN), Tsung-Teh Wu (Mayo Clinic, Rochester, MN), and Kathleen J. Yost (Mayo Clinic, Rochester, MN).

Conflicts of interest

These authors disclose the following: Alain M. Schoepfer received consulting fees and/or speaker fees and/or research grants from Adare Pharmaceuticals, Inc, AstraZeneca, AG, Switzerland, Aptalis Pharma, Inc, Dr. Falk Pharma, GmbH, Germany, Glaxo Smith Kline, AG, Nestlé S. A., Switzerland, Receptos, Inc, and Regeneron Pharmaceuticals, Inc. Ikuo Hirano received research grants from Shire plc, Regeneron Pharmaceuticals, Inc, Celgene Corp/Receptos, Inc, and Adare Pharmaceuticals, Inc and consulting fees from Adare Pharmaceuticals, Inc, Shire plc, Regeneron Pharmaceuticals, Inc, Allakos, and Celgene Corp./Receptos, Inc. Marcel Zwahlen received research grants from Adare Pharmaceuticals, Inc, AstraZeneca, AG Switzerland, Aptalis Pharma, Inc, Dr. Falk Pharma, GmbH, Germany, Glaxo Smith Kline, AG, Nestlé S. A., Switzerland, Receptos, Inc, and Regeneron Pharmaceuticals, Inc. Claudia E. Kuehni received research grants from Adare Pharmaceuticals, Inc, AstraZeneca, AG, Switzerland, Aptalis Pharma, Inc, Dr. Falk Pharma, GmbH, Germany, Glaxo Smith Kline, AG, Nestlé S. A., Switzerland, Receptos, Inc, and Regeneron Pharmaceuticals, Inc. Jeffrey A. Alexander received research grants and/or consulting fees from Merck & Co, Inc, Meritage Pharma, Inc, and Aptalis Pharma, Inc; receives royalties for commercial use of the Mayo Dysphagia Questionnaire - 30 Day; and also has financial interest in Meritage Pharma, Inc. Evan S. Dellon received research grants from Adare, Celgene Corp/Receptos, Inc, Meritage Pharma, Inc, Miraca, Nutricia, Regeneron, and Shire; has received consulting fees from Adare, Alivio, Allakos, AstraZeneca, AG, Banner, Enumeral, GSK, Celegene/Receptos, Inc, Regeneron Pharmaceuticals, Inc, and Shire; and has received educational grants from Banner and Holoclara. John Leung received research grants from Meritage Pharma, Inc. Margaret H. Collins received consulting fees from Adare Pharmaceuticals, Inc, Banner Life Sciences, Biogen Idec, Meritage Pharma, Inc, Novartis, AG, Receptos, Inc, Regeneron Pharmaceuticals, Inc, and Shire plc and has received contractual funds from Meritage Pharma, Inc, Receptos, Inc, and Regeneron Pharmaceuticals, Inc. Yvonne Romero collaborated on projects supported by Aptalis Pharma, Inc and Meritage Pharma, Inc and receives royalties for commercial use of the Mayo Dysphagia Questionnaire - 30 Day. Glenn T. Furuta has received consulting fees from Shire; is cofounder of EnteroTrack; receives royalties from UpToDate, Inc; and is also a founder of EnteroTrack, LLC. Sandeep K. Gupta received consulting fees and/or speaker fees from Abbott Laboratories, Nestlé S. A., QOL, Receptos, Inc, and Meritage Pharma, Inc. Seema S. Aceves received consulting fees from Receptos, Inc, and Regeneron Pharmaceuticals, Inc. Mirna Chehade has received consulting fees from Actelion, Shire, and Allakos; has received funding from Nutricia; and has received clinical trial funding from Regeneron and Shire. Jonathan M. Spergel has received consulting fees from DBV Technology, Regeneron Pharmaceuticals, Inc, and Shire Pharmaceuticals, Inc. Gary W. Falk receives research support from Celgene Corp, Regeneron Pharmaceuticals, Inc, Shire, and Adare Pharmaceuticals, Inc. Gail M. Comer received consulting fees from Aptalis Pharma, Inc and Adare Pharmaceuticals, Inc. Alex Straumann received consulting fees and/or speaker fees and/or research grants from Actelion, AG, Switzerland, AstraZeneca, AG, Switzerland, Adare Pharmaceuticals, Inc,

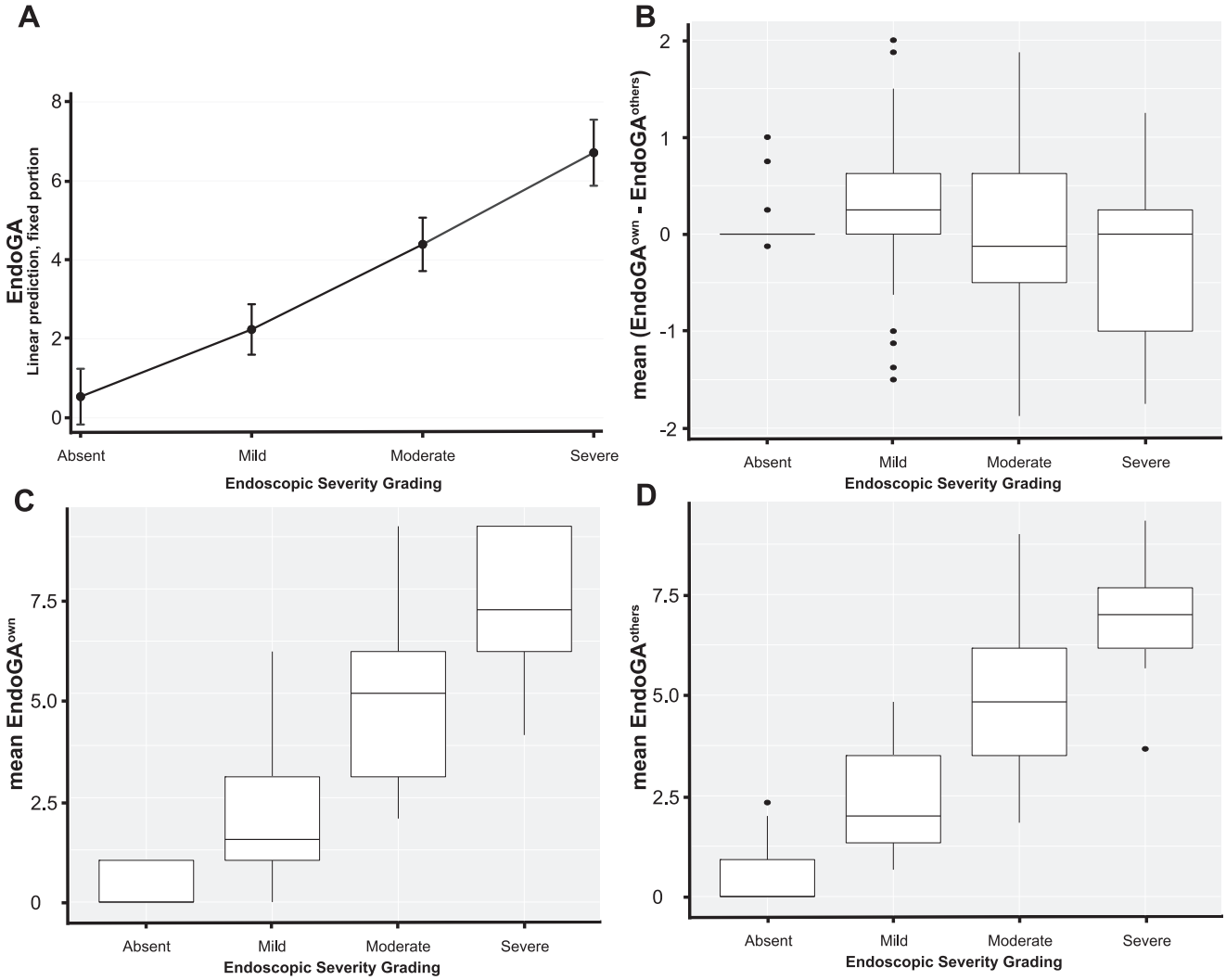
Dr. Falk Pharma, GmbH, Germany, Glaxo Smith Kline, AG, Nestlé S. A., Switzerland, Novartis, AG, Switzerland, Pfizer, AG, and Regeneron Pharmaceuticals, Inc. Ekaterina Safroneeva received consulting fees from Aptalis Pharma, Inc,/Celgene Corp, Novartis, AG, Switzerland, and Regeneron Pharmaceuticals, Inc. The remaining authors disclose no conflicts.

Disclaimer

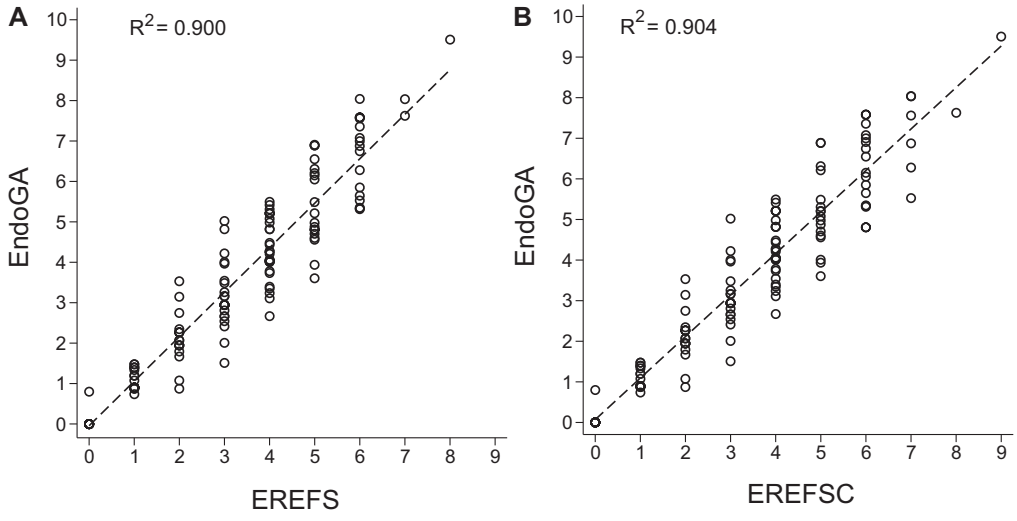
No granting agency/sponsor was involved in the study design, all the data analyses, interpretation of the data, and drafting of this paper. Whilst the data for score development and validation were collected by international Eosinophilic Esophagitis Activity Index study group (clinicaltrials.gov ID: NCT00939263 for EEsAI study), those for responsiveness were collected by Aptalis Pharma, Inc., as a part of their clinical study (clinicaltrials.gov ID: NCT 01386112 for trial).

Funding

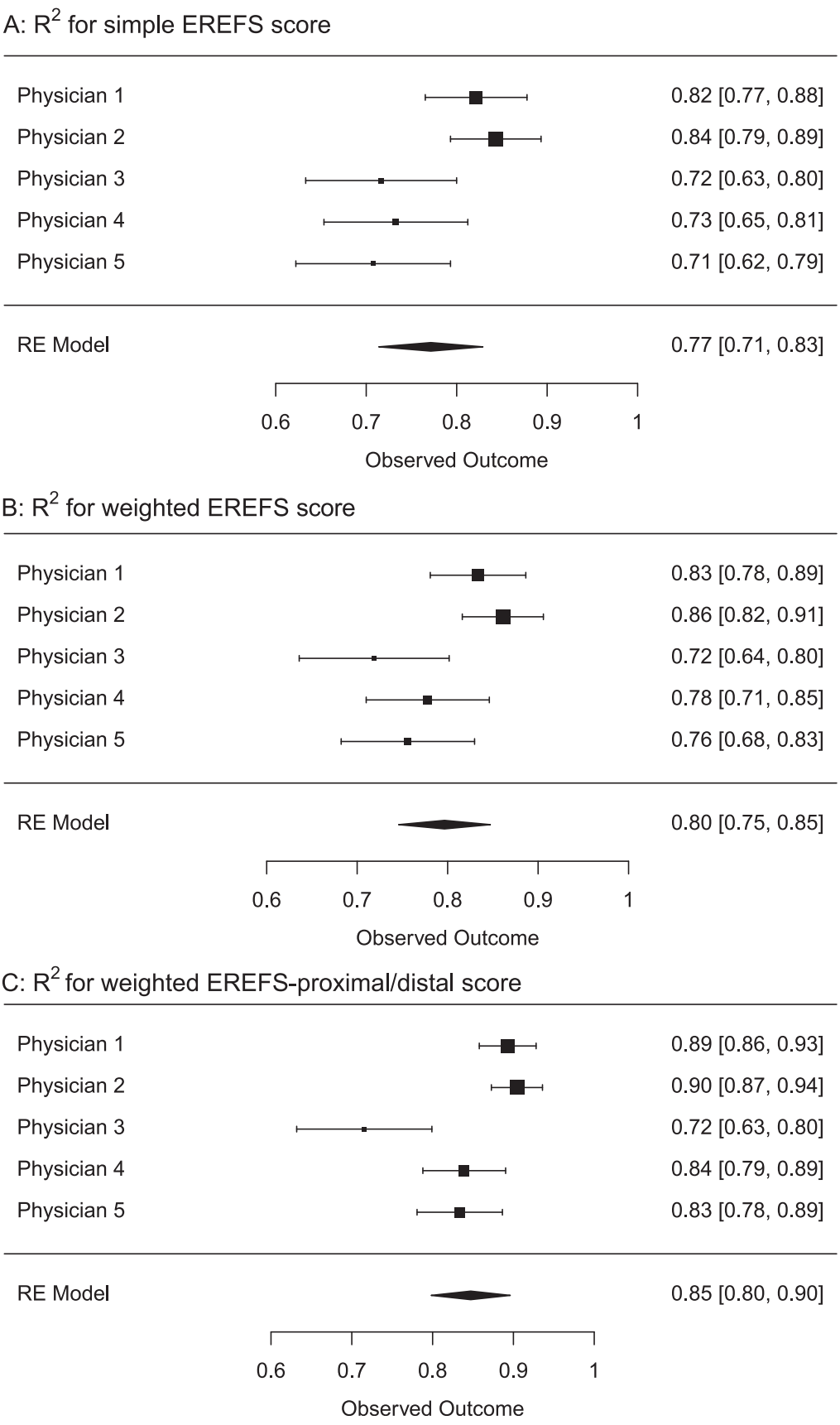
Supported by the following grants from the Swiss National Science Foundation (grant no. 32003B_135665/1 to A.M.S., A.S., C.K., and M.Z. and 32473B_160115/1 to A.M.S.), AstraZeneca AG, Switzerland, Adare Pharmaceuticals Inc, Dr. Falk Pharma GmbH, Germany, Glaxo Smith Kline AG, Nestlé S. A., Switzerland, Receptos Inc, Regeneron Pharmaceuticals Inc, and The International Gastrointestinal Eosinophil Researchers (TIGER). The data for score responsiveness were generated by Aptalis Pharma, Inc (currently owned by Adare Pharmaceuticals, Inc).



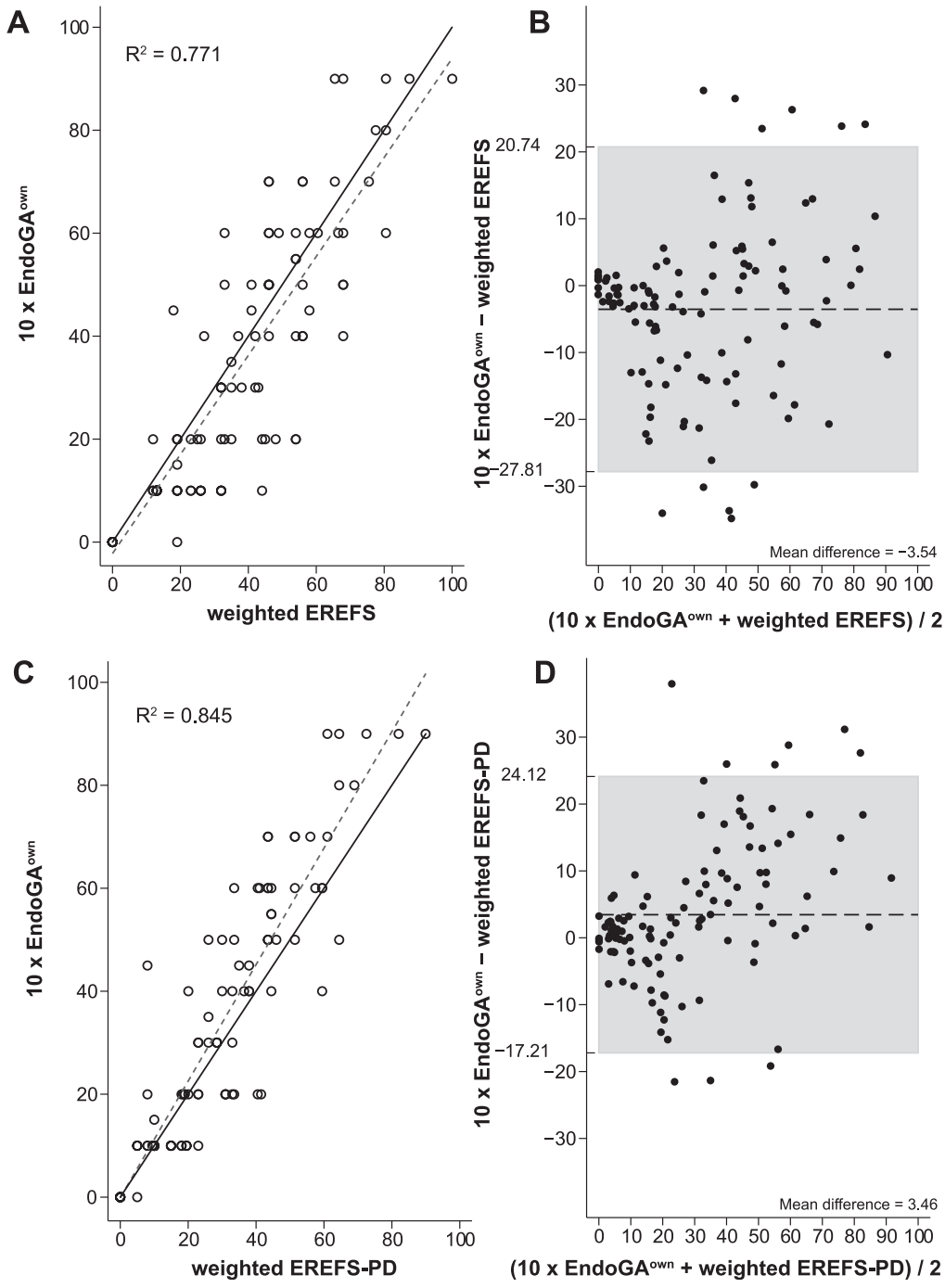
Supplementary Figure 1. Relationship between EndoGA and overall severity grading: (*A*) Relationship between fixed portion of estimated EndoGA and overall severity grading. *Vertical lines* indicate 95% confidence interval. *Box-and-whiskers plots* of (*B*) show mean difference between EndoGA for EGD performed in one's own center (EndoGA^{own}) and EndoGA for EGD performed in another center (EndoGA^{others}), (*C*) EndoGA^{own}, and (*D*) EndoGA^{others} versus overall severity grading. EndoGA, endoscopist global assessment.



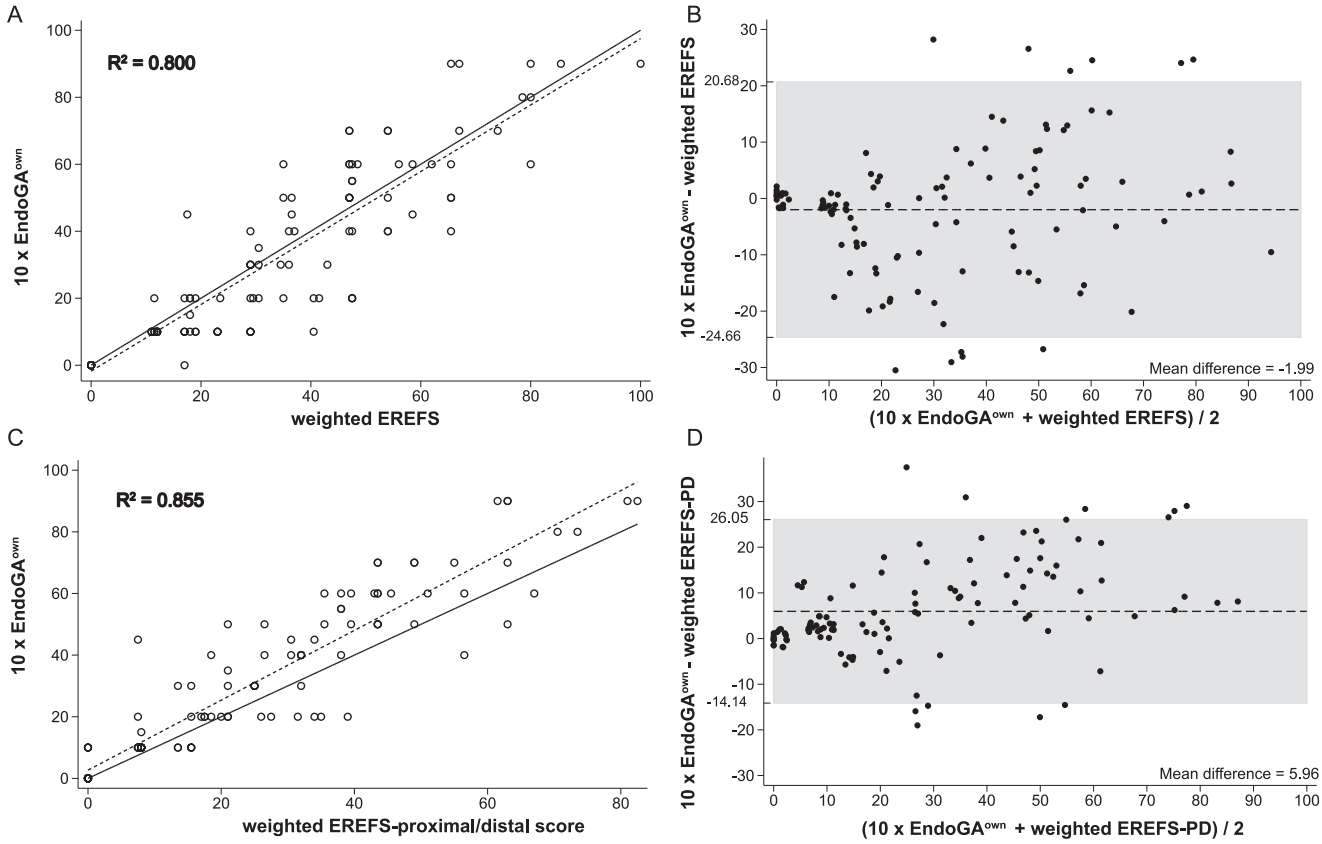
Supplementary Figure 2. Correlation between simple EREFS scores and EndoGA without crêpe-paper (*A*) and with crêpe-paper (*B*) in the evaluation group. EndoGA, endoscopist global assessment; EREFS, edema, rings, exudates, furrows, and strictures.



Supplementary Figure 3. The meta-analyzed R^2 values for (*A*) simple EREFS score, (*B*) weighted EREFS score, and (*C*) weighted EREFS-proximal/distal score. EREFS, edema, rings, exudates, furrows, and strictures.



Supplementary Figure 4. Calibration plots of EndoGA^{own} versus weighted scores in the validation group: (*A*) weighted EREFS score and (*C*) weighted EREFS-proximal/distal (EREFS-PD) score. *Solid line* has a slope of 1 and represents an ideal relationship between a perfect score and EndoGA values. *Dashed line* is the regression line fit to the data. The Bland-Altman plots for the agreement between the weighted scores and 10×EndoGA in the validation group: (*B*) weighted EREFS score and (*D*) weighted EREFS-PD score. *Gray box* indicates the 95% confidence intervals limits of agreement. EndoGA, endoscopist global assessment; EREFS, edema, rings, exudates, furrows, and strictures; PD, proximal/distal.



Supplementary Figure 5. Calibration plots of EndoGA^{own} versus for weighted scores developed on the basis of EndoGA^{own} in the validation group: (*A*) weighted EREFS score and (*C*) weighted EREFS-proximal/distal (EREFS-PD) score. *Solid line* has a slope of 1 and represents an ideal relationship between a perfect score and EndoGA values. *Dashed line* is the regression line fit to the data. The Bland-Altman plots for the agreement between the weighted scores and 10×EndoGA in the validation group: (*B*) weighted EREFS score and (*D*) weighted EREFS-PD score. *Gray box* indicates the 95% confidence intervals limits of agreement. EndoGA, endoscopist global assessment; EREFS, edema, rings, exudates, furrows, and strictures; PD, proximal/distal.

Supplementary Table 1. The EREFS Classification and Grading System–based Score

Feature	Grading	Definition	Points
Major features			
Rings (<1-cm length)	Grade 0	None	0
	Grade 1	Mild (subtle circumferential ridges)	1
	Grade 2	Moderate (distinct rings that do not impair passage of standard diagnostic adult endoscope [outer diameter, 8–10 mm])	2
	Grade 3	Severe (distinct rings that do not permit passage of diagnostic endoscope)	3
Exudates	Grade 0	None	0
	Grade 1	Mild (lesions involving <10% of esophageal surface area)	1
	Grade 2	Severe (lesions involving >10% of esophageal surface area)	2
Furrows	Grade 0	Absent	0
	Grade 1	Present	1
Edema	Grade 0	Absent (distinct vascularity present)	0
	Grade 1	Loss of clarity or absence of vascular markings	1
Stricture (≥1-cm length)	Grade 0	Absent	0
	Grade 1	Present	1
	OR	OR	
	Grade 0	None	NA
	Grade 1	Low-grade (esophageal diameter 11–13 mm, passage of standard endoscope possible against mild resistance)	NA
	Grade 2	Intermediate-grade (esophageal diameter 7–10 mm, passage of 6–mm outer diameter endoscope possible, but impossible with standard endoscope [8- to 10-mm outer diameter])	NA
	Grade 3	High-grade (passage of 6-mm outer diameter endoscope is not possible)	NA
Total score			8
Minor feature			
Crêpe-paper	Grade 0	Absent	0
	Grade 1	Present	1
Total score including crêpe paper			9

NOTE. In addition to absence/presence of strictures, we evaluated whether low-grade, intermediate-grade, or high-grade stricture(s) were found.
NA, not applicable.

Supplementary Table 2. Endoscopic Findings in Proximal and Distal Esophagus and in Esophagus Overall in All Patients and Evaluation and Validation Groups

Characteristic	Proximal		Distal		Overall	
	Frequency	%	Frequency	%	Frequency	%
Endoscopic findings (n = 266)						
Rings:						
Absent	104	39.1	81	30.5	74	27.8
Mild	89	33.5	99	37.2	99	37.2
Moderate	58	21.8	73	27.4	77	28.9
Severe	12	4.5	12	4.5	16	6.0
Missing	3	1.1	1	0.4	0	0.0
Strictures:						
Absent	223	83.8	192	72.2	178	66.9
Present	40	15.0	73	27.4	88	33.1
Missing	3	1.1	1	0.4	0	0.0
Exudates:						
Absent	205	77.1	183	68.8	177	66.5
Mild	46	17.3	67	25.2	73	27.4
Severe	14	5.3	14	5.3	16	6.0
Missing	1	0.4	2	0.8	0	0.0
Furrows:						
Absent	137	51.5	105	39.5	102	38.3
Present	128	48.1	160	60.2	164	61.7
Missing	1	0.4	1	0.4	0	0.0
Edema:						
Absent	140	52.6	114	42.9	113	42.5
Present	125	47.0	151	56.8	153	57.5
Missing	1	0.4	1	0.4	0	0.0
Crêpe-paper:						
Absent	239	89.8	239	89.8	234	88.0
Present	26	9.8	26	9.8	32	12.0
Missing	1	0.4	1	0.4	0	0.0
Endoscopic findings (evaluation group, n = 146)						
Rings:						
Absent	54	37.0	44	30.1	39	26.7
Mild	46	31.5	54	37.0	53	36.3
Moderate	36	24.7	40	27.4	44	30.1
Severe	8	5.5	7	4.8	10	6.8
Missing	2	1.4	1	0.7	0	0.0
Strictures:						
Absent	117	80.1	104	71.2	96	65.8
Present	27	18.5	41	28.1	50	34.2
Missing	2	1.4	1	0.7	0	0.0
Exudates:						
Absent	108	74.0	100	68.5	97	66.4
Mild	28	19.2	36	24.7	39	26.7
Severe	9	6.2	8	5.5	10	6.8
Missing	1	0.7	2	1.4	0	0.0
Furrows:						
Absent	62	42.5	50	34.2	49	33.6
Present	83	56.8	95	65.1	97	66.4
Missing	1	0.7	1	0.7	0	0.0
Edema:						
Absent	67	45.9	61	41.8	61	41.8
Present	78	53.4	84	57.5	85	58.2
Missing	1	0.7	1	0.7	0	0.0
Crêpe-paper:						
Absent	129	88.4	129	88.4	127	87.0
Present	16	11.0	16	11.0	19	13.0
Missing	1	0.7	1	0.7	0	0.0

Supplementary Table 2. Continued

Characteristic	Proximal		Distal		Overall	
	Frequency	%	Frequency	%	Frequency	%
Endoscopic findings (validation group, n = 120)						
Fixed rings:						
Absent	50	41.7	38	31.7	35	29.2
Mild	43	35.8	44	36.7	46	38.3
Moderate	22	18.3	33	27.5	33	27.5
Severe	4	3.3	5	4.2	6	5.0
Missing	1	0.8	0	0.0	0	0.0
Strictures:						
Absent	97	80.8	82	68.3	80	66.7
Present	18	15.0	32	26.7	34	28.3
Missing	5	4.2	6	5.0	6	5.0
Exudates:						
Absent	0	0.0	0	0.0	0	0.0
Mild	106	88.3	86	71.7	82	68.3
Severe	13	10.8	34	28.3	38	31.7
Missing	1	0.8	0	0.0	0	0.0
Furrows:						
Absent	75	62.5	55	45.8	53	44.2
Present	45	37.5	65	54.2	67	55.8
Missing	0	0.0	0	0.0	0	0.0
Edema:						
Absent	73	60.8	53	44.2	52	43.3
Present	47	39.2	67	55.8	68	56.7
Missing	0	0.0	0	0.0	0	0.0
Crêpe-paper:						
Absent	110	91.7	110	91.7	107	89.2
Present	10	8.3	10	8.3	13	10.8
Missing	0	0.0	0	0.0	0	0.0

Supplementary Table 3. Multivariable Linear Regression Model With Endoscopist Global Assessment as an Outcome and Endoscopic Features (Dependent Variables): One For Each Expert Endoscopist

	Expert 1			Expert 2			Expert 3			Expert 4			Expert 5		
	Coeff	95% CI	*P* value	Coeff	95% CI	*P* value	Coeff	95% CI	*P* value	Coeff	95% CI	*P* value	Coeff	95% CI	*P* value
Rings															
None	0.000	Ref.		0.000	Ref.		0.000	Ref.		0.000	Ref.		0.000	Ref.	
Mild	1.219	0.773–1.664	<.001	1.917	1.575–2.259	<.001	0.879	0.596–1.162	<.001	1.083	0.693–1.474	<.001	1.113	0.697–1.529	<.001
Moderate	3.049	2.516–3.583	<.001	2.755	2.345–3.164	<.001	1.683	1.344–2.023	<.001	1.614	1.146–2.082	<.001	1.576	1.077–2.075	<.001
Severe	3.672	2.600–4.743	<.001	3.147	2.325–3.969	<.001	2.694	2.013–3.375	<.001	1.883	0.943–2.822	<.001	1.722	0.721–2.723	<.001
Strictures															
Absent	0.000	Ref.		0.000	Ref.		0.000	Ref.		0.000	Ref.		0.000	Ref.	
Low-grade	0.532	0.069–0.994	.025	0.625	0.271–0.980	.001	1.578	1.284–1.872	<.001	0.749	0.347–1.151	<.001	0.591	0.163–1.020	.007
Intermediate/high	1.302	0.395–2.208	.005	0.962	0.267–1.658	.007	3.064	2.487–3.640	<.001	1.975	1.180–2.770	<.001	0.875	0.028–1.721	.043
Exudates															
None	0.000	Ref.		0.000	Ref.		0.000	Ref.		0.000	Ref.		0.000	Ref.	
Mild	1.593	1.171–2.015	<.001	0.479	0.155–0.803	.004	0.295	0.027–0.563	.031	2.097	1.729–2.466	<.001	2.284	1.892–2.676	<.001
Severe	3.586	2.885–4.288	<.001	2.470	1.932–3.008	<.001	0.638	0.191–1.084	.005	3.947	3.332–4.562	<.001	4.410	3.754–5.065	<.001
Furrows															
Absent	0.000	Ref.		0.000	Ref.		0.000	Ref.		0.000	Ref.		0.000	Ref.	
Present	0.878	0.413–1.344	<.001	1.100	0.741–1.455	<.001	0.511	0.215–0.807	.001	0.556	0.148–0.964	.008	0.515	0.080–0.949	.021
Edema															
Absent	0.000	Ref.		0.000	Ref.		0.000	Ref.		0.000	Ref.		0.000	Ref.	
Present	1.607	1.164–2.050	<.001	1.021	0.681–1.361	<.001	0.319	0.037–0.601	.027	1.165	0.776–1.553	<.001	1.287	0.873–1.701	<.001
Crêpe-paper															
Absent	0.000	Ref.		0.000	Ref.		0.000	Ref.		0.000	Ref.		0.000	Ref.	
Present	0.789	0.212–1.366	.008	–0.502	–0.493 to 0.392	.823	0.352	–0.014 to 0.720	.059	0.701	0.196–1.205	.007	0.930	0.393–1.468	.001
Constant	0.209	–0.157 to 0.574	.261	–0.037	–0.317 to 0.244	.796	–0.111	–0.343 to 0.122	.348	0.147	–0.174 to 0.467	.367	0.086	–0.256 to 0.427	.621
R^2	0.890			0.885			0.902			0.882			0.870		

NOTE: The constant represents the predicted Endoscopist Global Assessment value, when all values of independent variables are set to reference category. R^2 is a measure of the extent to which the regression model describes the observed data. The closer R^2 is to 1, the more precise the regression model is.
CI, confidence interval; Coeff, coefficient; Ref., reference.

Supplementary Table 4. Weighted EREFS Scores Based on Average of 5 values of EndoGA and Single EndoGA^{own} Values

	Weighted EREFS score based on average of 5 values of EndoGA (total set to 100)	New weighted EREFS score based on EndoGA^{own} (total set to 100)
Rings		
None	0	0
Mild	13	12
Moderate	23	19
Severe	35.5	33.5
Exudates		
None	0	0
Mild	14	18
Severe	33.5	38
Furrows		
Absent	0	0
Present	6	6
Edema		
Absent	0	0
Present	13	11
Strictures		
Absent	0	0
Present	12	11.5
Sum of the score	100	100

EndoGA, endoscopist global assessment; EREFS, edema, rings, exudates, furrows, and strictures.

Supplementary Table 5. Weighted EREFS-Proximal/Distal Scores Based on Average EndoGA and Single EndoGA^{own} Values

	Based on average value of 5 EndoGA values Score	New based on EndoGA^{own} Score
	(total set to 100)	(total set to 100)
Rings		
None	0	0
Mild in prox. and/or dist.	10	8
Moderate in prox. and/or dist.	18	13.5
Severe in prox. and/or dist.	18	24
Exudates		
None	0	0
Mild in prox. and/or dist.	15	18.5
Severe in prox. and/or dist.	19.5	24.5
Furrows	—	—
Absent		
Present in prox. and/or dist.		
Edema		—
Absent	0	
Present in prox. and/or dist.	5	
Crêpe-paper		
Absent	0	0
Present in prox. and/or dist.	5	6.5
Strictures		
Absent	0	0
Low-grade in prox. and/or dist.	8	7.5
Intermediate/high in prox. and/or dist.	18	14.5
Severe rings		
Absent in prox. and dist.	0	0
Present in prox. and dist.	8	–12
Severe exudates		
Absent in prox. and dist.	0	0
Present in prox. and dist.	13	13.5
Furrows		
Absent in prox. and dist.	0	0
Present in prox. and dist.	9	9.5
Edema		
Absent in prox. and dist.	0	0
Present in prox. and dist.	4.5	7.5
Crêpe-paper	—	—
Absent in prox. and dist.		
Present in prox. and dist.		
Sum of the score	100	100

dist., distal; EndoGA, endoscopist global assessment; EREFS, edema, rings, exudates, furrows, and strictures; prox., proximal.

Supplementary Table 6. Responsiveness of the Developed Scores

	Baseline to end of treatment slope of change in drug-treated group (converted from 0 to 100 for simple score)	Baseline to end of treatment slope of change in placebo-treated group (converted from 0 to 100 for simple score)	*P* value
Simple EREFS total score	–1.94 (–24.3)	–0.37 (–4.6)	.052
Simple EREFS inflammatory[a]	–1.62 (–20.3)	–0.37 (–4.6)	.04
Simple EREFS fibrotic[a]	–0.31 (3.9)	0.00 (0.0)	.62
Weighted EREFS total score	–23.5	–6.6	.12
Weighted EREFS inflammatory[a]	–19.9	–6.9	.13
Weighted EREFS fibrotic[a]	–3.6	0.4	.59
Weighted EREFS-proximal/distal score total score	–23.8	–8.4	.11
EREFS-proximal/distal score inflammatory[a]	–20.8	–10.6	.17
Weighted EREFS-proximal/distal score fibrotic[a]	–3.1	2.1	.39

EREFS, edema, rings, exudates, furrows, and strictures.

[a]In addition to the overall EREFS-based scores, we also separately examined the treatment effect on inflammatory endoscopic features (exudates, furrows, and edema) and fibrotic endoscopic features (rings, strictures, [and crêpe-paper for EREFS-proximal/distal score]).