Eosinophilic esophagitis: Comparison of clinical, endoscopic and histological scoring systems

Eosinophile Ösophagitis: Vergleich klinischer, endoskopischer und histologischer Scoringsysteme

Authors

Nikolas Johannes Lorenz1, Alexander Link1, Piotr Czapiewski1,2, Ulrike von Arnim1

Affiliations
1 Gastroenterologie, Hepatologie und Infektiologie, Universitätsklinikum Magdeburg, Magdeburg, Germany
2 Pathologie, Städtisches Klinikum Dessau, Dessau, Germany

Key words
Eosinophilic esophagitis, remission induction, remission maintenance, scoring systems

Schlüsselwörter
eosinophile Ösophagitis, Remissionsinduktion, Remissionserhaltung, Scoringsystem

received 19.09.2021
accepted 15.04.2022
published online 23.11.2022

Bibliography
Z Gastroenterol 2022; 60: 1779–1786
DOI 10.1055/a-1855-1974
ISSN 0044-2771
© 2022. The Author(s).
This is an open access article published by Thieme under the terms of the Creative Commons Attribution-NonDerivative-NonCommercial-License, permitting copying and reproduction so long as the original work is given appropriate credit. Contents may not be used for commercial purposes, or adapted, remixed, transformed or built upon. (https://creativecommons.org/licenses/by-nc-nd/4.0/).

Correspondence
Ulrike von Arnim
Universitätsklinikum Magdeburg AöR
Department of Gastroenterology, Hepatology and Infectious diseases, Leipzigzer Str. 44, 39120 Magdeburg, Germany
ulrike.vonarnim@med.ovgu.de

ABSTRACT

Background Eosinophilic Esophagitis (EoE) has received increasing attention as a disease entity, and it is now recognized as an important disorder of the Upper Gastrointestinal Tract. Topical corticosteroids (tCS) are effective in clinical-pathological remission induction (RI) and remission maintenance (RM) of active EoE. With scoring systems, such as clinical (SDI), endoscopic (EREFS), and histological (EoEHSS) systems, EoE can be graded, and its disease activity can be assessed.

Objective To discover how closely results within each of the three scoring systems SDI, EREFS, and EoEHSS are correlated between initial diagnosis (ID), RI, and RM, and to determine how well scores from the three systems are intercorrelated at each time point.

Methods Retrospective cohort analysis of patients with active EoE was performed between 2006 and 2020, with follow-up for up to 6 years. SDI, EREFS and EoEHSS scores were recorded at ID, at RI, and in RM. Evaluation employed descriptive statistics, the Friedman test, and Bonferroni-corrected post hoc pairwise comparisons.

Results At RI 29 and at RM 19 EoE patients provided data. Significant correlations were found between EREFS and EoEHSS at RI and in RM. Pairwise comparisons showed significant differences between ID and RI for SDI, for EREFS, and for EoEHSS.

Conclusion The scoring systems tested did not show intercorrelation at ID. Comparison revealed significant differences for SDI, EREFS, and EoEHSS between the systems at ID and RI, but not in RM, during tCS treatment. These results underline the efficacy of tCS (at RI and RM) in the treatment of active EoE.
Auswertung erfolgte mittels deskriptiver Statistik, Friedman-Test und Bonferroni-korrigierter Post-hoc-Vergleiche.

**Ergebnisse** Zum Zeitpunkt der RI konnten bei 29 sowie zur RE 19 EoE-Patienten analysiert werden. Alle EoE-Patienten wurden im Mittel über 13 Wochen bis zur RI bzw. im Mittel über 21 Monate mit tCS bis zur RE behandelt. Signifikante Korrelationen zeigte der EREFS zum EoEHSS bei RI sowie bei RE. Die Paarvergleiche mittels Bonferroni-korrigierter Post-hoc-Tests ergaben signifikante Unterschiede zwischen ED und RI für SDI, EREFS und EoEHSS.

**Schlussfolgerungen** Die Ergebnisse zeigen, dass bei ED die untersuchten Bewertungssysteme zur Bestimmung der Krankheitsaktivität nicht untereinander korrelieren. Der Vergleich der Bewertungssysteme erbringt signifikante Unterschiede zu den Zeitpunkten ED und RI für SDI, EREFS und EoEHSS, jedoch nicht in der RE unter tCS-Therapie. Diese Ergebnisse unterstreichen die Effektivität der tCS in der RI- und RE-Therapie der aktiven EoE.

---

**Introduction**

Eosinophilic esophagitis (EoE) is a chronic inflammatory disorder of the esophagus, mediated by type 2 T-helper cells. It is manifested clinically by symptoms of esophageal dysfunction and histologically by eosinophil-dominant infiltration of the esophageal mucosa. Reliable diagnosis requires the exclusion of other possible causes of esophageal eosinophilic infiltration[1, 2]. To objectify the clinical symptoms, various scoring systems are available, some validated [3, 4, 5] and some not; the latter include the Straumann Dysphagia Instrument (SDI) [6]. To take account of the highly variable picture obtained by endoscopy, endoscopic diagnosis can be standardized by use of the Endoscopic Reference Score (EREFS), which shows moderate, but adequate, inter- and intra-observer consistency [7, 8]. From the histopathological side, the Eosinophilic Esophagitis Histology Scoring System (EoEHSS) serves as a further instrument of assessment, allowing classification of the extent and severity of histologically determined EoE activity. Furthermore, the EoEHSS allows distinction between untreated and treated EoE patients [9], and it shows high inter-rater and intra-rater consistency [10]. Topically administered corticosteroids (tCSs) are efficacious in both the induction [11, 12, 13] and the maintenance [14, 15, 16, 17] of clinical-pathological remission of EoE. Several studies have shown that the clinical symptoms and the histological activity of EoE are frequently only poorly correlated with one another [18, 19, 20]. In this study we first investigated, separately for each of the three assessment systems SDI, EREFS, and EoEHSS, the extent to which results at the times of initial diagnosis (ID), induction of remission (RI), and maintenance of remission (RM) were correlated (see ▶ Fig. 1A). We then investigated, for each of these time points, how the results from ID, RI, and RM were correlated with one another (▶ Fig. 1B).

**Materials and methods**

**Study design**

This monocentric, retrospective observational study was conducted at the Magdeburg University Hospital over a period from 2006 to 2020. We investigated a cohort of EoE patients, who after ID were treated with tCSs and subsequently achieved RI and then RM, by conducting assessments at the times mentioned. The cohort analysis was approved by the Ethics Committee of Magdeburg University Hospital (AZ R18–18).

EoE was defined according to the current consensus recommendations [1, 2]. Symptoms of esophageal dysfunction and of eosinophilic esophageal infiltration (≥ 15 eosinophil granulocytes per high-power field, EoS/HPF) in at least one out of a maximum of six esophagus biopsies after exclusion of other possible causes of esophageal eosinophilia. Non-adherence to the therapy, changes in diet or lifestyle, inadequately assessable follow-up examinations, or fewer than two evaluable Oesophageo-Gastro-Duodenoscopies (OGDs) likewise resulted in exclusion. Earlier endoscopic interventions such as dilations, or incidental findings of axial hernias, were accepted. Patients of either sex and any age and ethnicity could be included. Clinical-histological remission was defined as a decrease in SDI by ≥ 3 points accompanied by eosinophilic infiltration of the esophageal mucosa of ≤ 15 EoS/HPF.

**Data acquisition**

Demographic data, symptom frequency and severity, any allergic comorbidities, conspicuous endoscopic findings, and relevant disease characteristics for the scoring systems under study were recorded.
Clinical score (SDI)

The SDI [6] was used for assessment of dysphagia. It quantifies and sums the severity and the frequency of dysphagic events, as follows. Frequency: “none” = 0, “once per week” = 1, “several times per week” = 2, “once per day” = 3, “several times per day” = 4. Severity: “swallowing unhindered” = 0, “slight sensation of resistance” = 1, “slight retching with delayed passage” = 2, “short period of obstruction necessitating intervention (e.g., drinking, breathing)” = 3, “longer-lasting period obstruction only removable by vomiting” = 4, “long-lasting complete obstruction requiring endoscopic intervention” = 5. The total possible score is thus 0–9.

Endoscopic score (EREFS)

Data were recorded retrospectively by reference to 60 complete digital images (56 statical, 4 videos) of the esophageal sections (proximal, middle, and distal), 17 uncomplete digital images and to the acquired verbal descriptions of the OGDs. Results were classified according to the validated) modified EREFS criteria [7]: “fixed rings” = 0–3, “strictures” = 0–1, “reddish furrows” = 0–1, “off-white exudates” = 0–2, “mucous membrane oedema” = 0–1, “crepe paper esophagus” = 0–1. The total possible score is thus 0–9.

Histological score (EoEHSS)

Histological assessment represents a central feature in the diagnosis of EoE [19]. The uniformly standardized evaluation of histological sections was performed according to the validated EoEHSS [9]. The graduation was performed in a standardized manner by staining the microscopic preparations with haematoxylin–eosin, and additionally with Giemsa stain, from distal through middle to proximal. For each esophageal section one pair was used for each location: two for proximal, two for central, and two for distal (in total six biopsies).

Severity (grade) and extent (stage) are each assessed by the use of eight histological features, each on a scale from 0 to 3, for each section of the esophagus (proximal, central, distal) separately. The total possible respective score, reflecting the greatest possible histopathological alteration, is thus 24 points for each esophageal section. If a feature cannot be assessed in staging biopsy, then the maximum total score is automatically reduced, independently of the other features, by 3 points. To take account of missing features, the score is expressed as a ratio: the score actually obtained is divided by the greatest possible score for the biopsy in question; for example, if all features are assessable and each has the highest possible value, then the result for severity and extent is 24/24 = 1.00, while if only seven of the eight features can be assessed and the total score is 12, then the result is 12/21 = 0.57.

The mean result of a biopsy, for each section of the esophagus, is then calculated as an average of extent score $E$ and severity score $S$, i.e., $(E + S)/2$. The mean total score for all three locations is similarly: \[ \frac{E_{\text{proximal}} + E_{\text{central}} + E_{\text{distal}}}{3} + \frac{S_{\text{proximal}} + S_{\text{central}} + S_{\text{distal}}}{3}. \]

Thus, the total overall score lies in the range 0–1, with 1 as the worst possible value.

Statistical analysis

All calculations were performed with GraphPad PRISM (version 7.05). Descriptive statistics were used to characterize the cohorts. For each patient and each point in time (ID, RI and RM) the scores for SDI, EREFS and EoEHSS were determined. The statistical significance of differences for the ordinarily scaled scoring systems were determined for pairs of samples by using the Wilcoxon pair-difference test. For groups of more than two samples the Friedman test was used and a two-factor repeated-measures analysis of variance (ANOVA) was conducted. Post hoc testing with the Dunn–Bonferroni correction and subsequent calculation of the size of the effect was performed. The correlations were assessed by means of Spearman’s coefficient ($r_s$): $r_s = 0.1$ poor correlation, $r_s = 0.3$ middling correlation, $r_s = 0.5$ strong correlation. The significance level was set to $p = 0.05$.

Results

Patient characteristics

Sixty EoE patients were included according to the in- and exclusion criteria (see Fig. 2). Of these, 29 (48 %) were assessable for two time points (ID and RI) and 19 (32 %) for all three time points (ID, RI and RM). Demographic and disease data for the 29 assessable patients are shown in Table 1. Previously, none of them had responded to a 6–8 week high-dose proton pump inhibitors (PPI) treatment to rule out PPI-responsive esophagitis. On average, these patients were treated for 13 weeks with TCS in their remission-inducing therapy; 19 (65 %) received budesonide-containing orodispersible tablets (BUD-SKT), 8 (28 %) a budesonide suspension (BUD-S), and 2 (7 %) oral fluticasone (sFLU). As 10 of the patients did not provide follow-up data (i.e., for remission), there remained 19 complete patient-data sets for assessment in RM. The mean duration of RM with TCS was 21 months; 15 (79 %) of the 19 patients were treated with BUD-SKT, 1 (5 %) with BUD-S, and 3 (16 %) with sFLU.

![Fig. 2 Flow diagram for in-/exclusion of patients.](image-url)
SDI at time points ID, RI and RM

Mean SDI (± standard deviation, SD) at ID was 5.8 ± 1.1. It decreased with statistical significance at RI to 1.6 ± 1.2 (N = 29, p < 0.001; shown in Fig. 3A). Dunn–Bonferroni tests conducted post hoc at RM gave a consistently low SDI of 0.4 ± 0.8 (N = 19, p = 0.144; shown in Fig. 3B).

EREFS at time points ID, RI, and RM

In all, 77 sets of endoscopic findings for ID (29), RI (29), and RM (19) were assessed. The EREFS at ID was 4.6 ± 1.7, and at RI it fell with statistical significance to 3.2 ± 2.0 (N = 29, p < 0.001; shown in Fig. 3A). Dunn–Bonferroni-corrected pairwise comparisons, performed subsequently, showed a consistently lower total EREFS score of 1.8 ± 1.8 (N = 19, p = 0.186; shown in Fig. 3B).

EoEHSS at time points ID, RI, and RM

At ID, a total of 174 biopsies from the proximal, central, and distal locations were taken, at RI 174 and at RM 114. Of these, 168 (97 %, from 29 patients) at ID, 172 (99 %, from 29 patients) at RI, and 112 (98 %, from 19 patients) at RM were evaluable. The EoEHSS was 0.41 ± 0.21 at ID and at RI it had decreased with statistical significance to 0.16 ± 0.16 (N = 29, p < 0.001; shown in Fig. 3A). Dunn–Bonferroni-corrected pairwise comparisons, performed subsequently, showed sustained low EoEHSS total scores at RM: 0.10 ± 0.10 (N = 19, p = 0.223; shown in Fig. 3B).

Clinical-histological remission

Clinical remission was attained (time point RI) on average after 13 weeks for 24 (83 %) of the 29 EoE patients; 23 (79 %) of these were in histological remission. With a combined criterion for clinical-histological remission, 18 (62 %) patients were identified as being in remission at RI. At time point RM, all the patients were complaint-free and had on average been in clinical remission for 21 months after their treatment. Two patients showed histological recurrence during treatment. Five patients were late responders who only attained histological remission at time point RM, so that – after discounting the patients with missing follow-up examinations after remission-maintenance tCS treatment – histological remission was observed in 17 (90 %) of the 19 patients assessed at RM. Clinical-histological remission at RM was found for 17 (90 %; not the same 17 as above) of these patients.

Mutual correlations between the ID, RI and RM total scores

When the scores at ID and RI were compared for the 29 EoE patients, no correlation was found. When the time points RI and RM were compared, strong correlation was found between EREFS and EoEHSS: at RI, r = 0.56, p < 0.05 (Fig. 4A) and at RM r = 0.61, p < 0.05 (Fig. 4B).

Discussion

In this study we investigated the relationships within and between clinical, endoscopic, and histological assessment systems in the
short and long term. Clinically significant results on the efficacy of tCSs, using the SDI after RI, have been demonstrated in placebo-controlled studies, which showed reductions from 2.7 [12] to 3.39 [6] score points. Our results support these, and in our study the SDI score was even more strongly reduced (by 4.2 points). Regarding the development of the SDI score in maintenance therapy, available data are sparse [21]. Greuter et al. have shown that among 23 EoE patients the mean SDI score decreased by 5.8 points from ID to RM [15]. Our own results confirmed this, with a mean decrease in SDI score of 5.4 points. The EREFS – as an important, established and validated endoscopic scoring system – has often been used as a surrogate marker for determining the usefulness of endoscopy in EoE patients [22, 23, 24], and it has inter alia been employed as a secondary endpoint for assessing the efficacy of modern biologicals in treatment of active EoE [25].

Significant differences between endoscopic assessment systems have been observed. The patient cohorts examined have included: children in short- and long-term therapy [14], and adults after RI following treatment with tCSs [13] and/or dietetic therapy [26] and after RM [17]. Because of separate recording of distal and proximal EREFS results [27, 28], and because of omission [29, 30, 31] or alteration [26, 32] of the assessment criteria, study results are difficult to compare. Our investigation showed a significant improvement in EREFS after RI for adults, with scores remaining constant during the maintenance phase. The results of a randomized, placebo-controlled study at RI [13] and of another prospective observational study at RM [15] gave comparable results in respect of improvement in EREFS after medication-based induction or maintenance of remission. In contrast to this situation, for EoEHSS only a few studies have been reported [28, 32, 33]. Frequently, the peak eosinophil count (PEC) has been at the centre of histological investigations in adult [29, 31] and child [34] EoE patients. This results from the current consensus guideline recommendation, according to which eosinophil count is a
determinative factor in diagnosis [2]. In a placebo-controlled study with oral BUD-S, Collins et al. found, after therapy, improvements not only in PEC [27], but also (in histological assessment of extent) in eosinophilic inflammation (EI), epithelial basal zone (BZH), eosinophilic abscesses (EA), eosinophilic surface layering (SL), dilated intercellular spaces and lamina propria fibrosis and (in histological assessment of severity) above all in EI, BZH, EA, SL and surface epithelial alteration [28]. In our study the total EoEHSS score decreased significantly, by an average of 0.25 points, after induction; in the subsequent maintenance therapy it decreased by 0.06 points. Our results show for the first time the course of change in EoEHSS during and following the induction of remission. Alongside the clinical and endoscopic scoring systems, EoEHSS also showed consistently low, and improved, scores in the maintenance phase.

Apart from long-term reduction in the various scores, the relationships between the scores are also decisive for prognosticating disease activity. The predictive power of the EREFS has been the subject of controversy. A prospective study has indicated that the diagnosis of EoE and the histological response to therapy can reliably be predicted on the basis of EREFS [26]. However, other studies have shown that EREFS only allows an inadequate clinical and histological prediction of disease activity [29, 31]. Our study showed that at ID none of the scores were intercorrelated. A randomized, placebo-controlled Phase 2 study with dupilumab confirmed the lack of correlation among SDI, EREFS, and EoEHSS for adults before the inception of therapy [35]. In an earlier prospective observation study [20], it was shown that symptoms could not reliably be used as clinical variables for assessing endoscopic and/or histological disease activity. A retrospective long-term study confirmed the lack of correlation between the severity of symptoms and the eosinophil count [36]. In our study EREFS and EoEHSS, after remission-inducing therapy with tCSs, were correlated with one another at RI (shown in Fig. 4A) and at RM (shown in Fig. 4B). SDI showed no correlation with EREFS and EoEHSS. In a double-blind, placebo-controlled Phase 2 study, changes in EoEHSS were weakly to moderately correlated with changes in EREFS after 12 weeks of treatment with BUD-S [28], in agreement with our own results. Furthermore, after treatment with dupilumab, correlation between these two scores was observed [33]. A retrospective paediatric study [32] showed weak ($r_s = 0.42, p < 0.001$) or very weak ($r_s = 0.24, p < 0.020$) correlations between EREFS and EoEHSS in both active and inactive EoE.

Ideally, validated scoring systems and large cohort sizes should be used to look for possible correlations with a view to predicting disease activity and treatment status, and to allow an estimation of the predictive power of such correlations. A limitation of our study is the relatively small sample size, although this is not atypical of EoE studies [6, 15, 21]. However, owing to the relatively strict inclusion criteria (see above, Study Design) our results retain their validity. Moreover, we used a retrospective study design in order to obtain clinical, endoscopic, and histological findings from EoE patients; this may have led to some distortion of results through selection bias. Some of the study data were acquired before the publication of the clinical [6], endoscopic [7], and histopathological [9] scoring systems, so that distortions could have arisen through incorrect classification (information bias). Our most recent data set from 2006 was acquired before the introduction of SDI (in 2010) and also before DSQ [3] and EEsAI PRO [5] were established (in 2013 and 2014 respectively). Our clinical analysis was made in detail and was documented; it was aimed specifically at recording the intensity and frequency of dysphagia, at discerning possibilities in the daily course of the patients’ disorder, and at determining the types of meal that gave rise to dysphagic events. This enabled us to determine these retrospectively from our data sets before the introduction of SDI. Both DSQ and EEsAI are less value in everyday practice, i.e., outside controlled clinical studies, on account of their complexity. Both of these assessment instruments need to be applied almost daily over a period of 14 [37] or respectively 7 days [5]. Moreover, EEsAI is recommended mainly for prospective studies [5].

Nonetheless, the strengths of our study lie in the long observation and treatment periods covered during RM, which alongside the clinical assessment was assessed by EREFS and EoEHSS, both of which were developed and validated specially for EoE. The clinical and endoscopic scoring was performed by an EoE expert (U.v.A.) and the data for them were acquired under her supervision. EoEHSS was assessed by an experienced pathologist (P.C.). According to current consensus conditions, a histological remission is defined as one showing < 15 Eos/HPF. In the literature, the use of stricter definitions has been reported; nonetheless, the limit values reported in this paper have generally been found.

Regarding clinical-histological remission, reports are not in uniform agreement. To the best of our knowledge, no long-term results, based on a combination of clinical and histological assessment tools, have so far been published. We therefore use in this work the definition of clinical-histological remission as used in a published study and provide one of the first reports of results of clinical-histological remission in the short and especially in the long term [13].

Our results show that a tCS-based induction therapy is – both clinically and histologically – highly effective and that this efficacy can be maintained during a long-term tCS therapy. This result demonstrates that adult EoE patients profit from tCS in the long term, as also found by Greuter et al. [30]. A further important finding in this study is that symptoms and histological activity in the maintenance phase are not intercorrelated. This has consequences for the surveillance of EoE patients; a regular observation of histological activity during maintenance therapy is indicated, in the interest of avoiding long-term complications.

In summary, our study has shown that results from the clinical, endoscopic and histological scoring systems were not intercorrelated at ID. They differed significantly after induction of remission, but not in therapy maintenance. During RM all the systems showed stable, low scores. These results underlie the efficacy of tCSs in both the short-term and the long-term treatment of active EoE.

Conflict of Interest

The authors declare that they have no conflict of interest.