EPIDERMOLYSIS BULLOSA OROPHARYNGEAL SEVERITY SCORE (EBOS): A MULTICENTRIC DEVELOPMENT AND RELIABILITY ASSESSMENT
Funding sources: Funding for this work was provided by Department of Dermatology, Stanford University, Stanford, CA, USA, by the Office of Research and Development, Palo Alto Veteran’s Affairs Medical Center, Palo Alto, CA, USA (Dr. Marinkovich), by D.eb.R.A. Mexico Foundation, Monterrey, Mexico, and by the Oral Medicine Unit, Department of Odontostomatological and Maxillofacial Sciences, Federico II University of Naples, Naples, Italy.

Conflict of interest: The authors have no conflict of interest to declare

Contents of the manuscript have not been previously published or presented in a congress and are not currently submitted elsewhere

BULLETED STATEMENT

What's already known about this topic?

• Children with any of type of epidermolysis bullosa (EB) are at higher risk of developing oropharyngeal lesions, involving either hard or soft tissues.

• Very few reports have been published aimed at developing an EB scoring system, giving little, if any, weight to the oropharyngeal component.

What does this study add?

• This study provides a new, objective, easy to perform, and reproducible scoring system for the oropharyngeal component in EB children, which has demonstrated an excellent inter- and intra-observer reliability.
ABSTRACT

**Background:** Epidermolysis bullosa (EB) is a rare genetic disorder characterized by constant formation of mucocutaneous blisters upon trivial trauma. All four EB types may show oropharyngeal lesions involving either hard or soft tissues. Currently, there are very few data on EB scoring including the oropharyngeal cavity.

**Objectives:** To set up an oropharyngeal severity score that was objective, valid, reliable, reproducible, easy to perform, and appropriate for all EB types.

**Methods:** In this multicentric study, oral medicine specialists developed a new score, the Epidermolysis Bullosa Oropharyngeal Score (EBOS). This measured oropharyngeal disease activity (erythema, atrophy, blisters, erosion/ulceration) and structural damage (microstomia, ankyloglossia, scarring phenotype beyond microstomia and ankyloglossia, enamel hypoplasia). It was tested on 92 patients with different types/subtypes of EB, and inter- and intra-observer reliability were assessed.

**Results:** The EBOS mean total score was 12.9±10.9 (range 0–33.5). Both inter-and intra-observer reliability for total score on all EB patients were considered excellent (ICC: 0.94; 95% confidence interval (CI): 0.90–0.96 and ICC: 0.90; 95%CI: 0.84–0.94, respectively). Even analyzing each single parameter of the disease activity and structural damage, a substantial-excellent correlation was found in the inter-observer (except for four sites) and intra-observer reliability. A significant correlation was found between EB types/subtypes and the EBOS median score ($p< .001$), but not between age and the EBOS mean total score in each group.

**Conclusions:** The EBOS score seems to represent an instrument capable to truly quantify the oropharyngeal severity in different types/subtypes of EB, demonstrating an excellent inter- and intra-observer reliability.
INTRODUCTION

Epidermolysis bullosa (EB) encompasses a group of inherited mucocutaneous disorders characterized by the occurrence of blisters onto the skin and mucous membranes following mild mechanical trauma.\(^1\) Recently, EB has been classified in four major types, based on the split at the ultrastructural level: intraepidermal or epidermolytic (EB simplex [EBS]), intra-lamina lucida or lucidolytic (Junctional EB [JEB]), sub-lamina densa or dermolytic (Dystrophic EB [DEB]), and mixed (Kindler syndrome), and about 30 different subtypes.\(^2\)

Many EB patients may suffer from systemic complications and mucosal lesions, such as genital, ocular, oropharyngeal ones.\(^3\)\(^-\)\(^5\) These may affect both hard and soft tissues, showing different features and degrees of severity.\(^6\) In all EB types oral soft tissue are fragile, resulting in frequent blister and/or erosion formation, accompanied, in some EB subtypes, by a scarring phenotype and, although rarely, oral milia.\(^7\)\(^-\)\(^10\) Similarly, oral hard tissues may show either a marked developmentally compromised enamel or minor structural defects with areas of surface pitting and furrowing.\(^11\)\(^-\)\(^15\)

Although until now no consensus statement on EB severity score has been established, two reports have been published aimed at developing an EB scoring system.\(^16\),\(^17\) These attempted to develop a method of scoring EB severity, evaluating too many variables (skin, height, weight, mucous membranes, nutritional status, cancer, etc), and giving little, if any, weight to the oropharyngeal component. Hence, the need to develop an independent mucosal scoring system, because a “single all-inclusive” score might be inadequate, as unable to truly reflect the severity of clinical conditions and correlation with disease prognosis. Our concern is that the oropharyngeal involvement has been previously\(^16\),\(^17\) evaluated solely based on subjective clinician's observations and patient’s reports (presence/absence of blisters/erosions over an “undefined” period of time, e.g., always, several per month,\(^7\) occasional, frequent, persistent\(^8\)). Consequently, this should render these scoring systems less
reliable and reproducible, causing the notion of “score” to collapse.

Considering the importance and impact of the oropharyngeal component on EB patients’ global health, we have developed a new separate score, called Epidermolysis Bullosa Oropharyngeal Severity Score (EBOS). Our purpose was to quantify and monitor the oropharyngeal involvement with an instrument that was as much as possible: (i) objective, in order to guarantee an effective and practical report of clinical signs far from subjective patient’s perception; (ii) valid, with content validity evaluated by experts in the field of oral medicine; (iii) reproducible and reliable worldwide among the same and different oral health care providers (oral medicine specialists, dermatologists, otorhinolaryngologists, paediatricians); (iv) easy to perform, so that to calculate the total score very quickly and, then, acceptable for patients, and (v) appropriate for all EB types/subtypes.
MATERIALS AND METHODS

Study design and Patients

This was a multicentric study collecting data from 92 EB patients between September 2010 and September 2011 coming from the EB Clinic at Lucile Packard Children’s Hospital, the adult Bullous Disease Clinic, Stanford, California (USA), and Dermatology Clinic at Istituto Tecnologico y de Estudios Superiores and D.eb.RA. Mexico Foundation, Monterrey (Mexico). The Department of Orofacial Sciences, School of Dentistry, University of California, San Francisco (USA), and the Oral Medicine Unit, Department of Odontostomatological and Maxillofacial Science, Federico II University of Naples, Naples (Italy) cooperated with them. All patients provided their written informed consent. This study was approved by the Ethical Committees of Stanford University and Instituto Tecnologico in Monterrey.

All patients were enrolled based on the following inclusion criteria:

1. Patients of both gender, all age and race, with the presence of typical mucocutaneous lesions of any EB type/subtype, as previously reported.\(^2\)

2. Diagnosis of EB based on skin biopsy with a routine histology and immunofluorescence antigen mapping (IFM), and/or, whereas available, electron microscopy (EM) and/or DNA analysis.

3. Patients able to give consent if older than 18 years. For minor patients consent was given by their parents or guardian.

At the time of admission exclusion criteria encompass:

1. Patients who had used topical corticosteroids and/or topical and/or systemic antifungal therapy during the previous 3 weeks, as capable to substantially modify the oropharyngeal clinical appearance.
2. Patients with present or past history of oropharyngeal malignancy and/or potentially malignant disorders.

**Generation and refinement of scoring items**

In origin, the EBOS included only the number of sites involved, plus microstomia, ankyloglossia, and enamel hypoplasia. After an accurate revision of the literature and discussion among authors, this score was abandoned as considered not really indicative of disease severity. Content validity was accurately revised and the EBOS was refined by introducing the nature of oropharyngeal lesions and the presence of a scarring phenotype in other parts of the oral cavity, beyond microstomia and ankyloglossia.

Eventually, a more appropriate EBOS was re-designed, including 2 different scores: disease activity and structural damage (Figure 1).

The first evaluated only clinical signs, as objective findings. Specifically, four were identified as key features of disease activity: erythema, atrophy, blister, erosion/ulceration. These signs were not scored in terms of quality (mild, moderate, severe), because too subjective among physicians, or quantity (number of lesions), because too difficult and confusing to calculate, as usually oropharyngeal lesions tend to be confluent. Clinically, atrophy in EB appeared similar to that seen in progressive systemic sclerosis, i.e., with vestibule obliteration, depapillated tongue, disappearance of palate rugae, blanching of buccal mucosa and/or soft palate.

The second evaluated the presence or absence of four parameters: microstomia, ankyloglossia, presence of intraoral scars beyond microstomia and ankyloglossia, such as vestibule obliteration, and enamel hypoplasia. These clinical features were considered more permanent, as a part of a previous damage and, therefore, not necessarily reflecting current disease activity. Microstomia was evaluated with the maximal mouth opening by measuring the distance from the marginal edge of the central
upper to lower incisors, along the inter-incisal line. In case of missing teeth in one or both jaws, measurement was done considering the distance between edentulous alveolar ridges, passing through the two craniomethic points: nasion and gnathion. A patient with a maximal aperture less than 35mm was considered as having microstomia, as previously reported.\textsuperscript{18} Ankyloglossia was evaluated by the ability of each patient to protrude his tongue over the lower incisors or edentulous alveolar ridge, move it over to the left and right side, and reach the premaxilla with the tip of the tongue. A patient unable to perform at least two of the above-mentioned movements was considered as having ankyloglossia.

The EBOS score

In order to increase the score sensitivity, the oropharyngeal cavity was divided in 13 different anatomic sites. The disease activity score evaluated each site affected by one or more clinical signs. We decided to assign 1 point to each clinical sign present in each anatomical site, leading to a total score ranging from 0 to 52. Conversely, the structural damage score evaluated the presence or absence of the 4 structural damages, assigning 2 point each to a total score ranging from 0 to 8 (Figure 1).

Grading system was based on the sum of both scores, reaching a final total score ranging from 0 to 60, rather than on virtually impossible task of determining accurately the percentage of each site involved by each type of lesion.

Inter- and Intra-observer reliability

Inter-observer reliability was evaluated in all patients, who were scored independently by two different physicians on the same day. Conversely, intra-observer reliability was assessed on a randomly selected group of patients (34 out of 92). Such patients were asked to come back after three hours and not to eat anything, use any kind of topical/mouthwash medication, drink alcohol or smoke cigarettes. In order to minimize recall bias, during the three hours of interval, the scorers were asked to see 100
consecutive pictures of patients with oropharyngeal blistering diseases, such as pemphigus vulgaris, mucous membrane pemphigoid, erosive lichen planus. Eventually, each patient was seen twice by the same physicians, and, on the second round, in a different and random order compared to his/her first visit. Time for scoring was also recorded.

**Statistic analysis**

Descriptive statistics of demographic characteristics and EB type/subtype distribution was calculated as mean ± standard deviation. The EBOS score was calculated as a mean value of the scores from two investigators for all 92 patients and the subgroup of 34 patients. Means, medians and interquartile ranges (IQR) in each EB type/subtype and each oropharyngeal site were also calculated. Inter- and intra-observer reliability for disease activity and structural damage (separately and grouped) were calculated by intraclass correlation coefficient (ICC) along with 95% confidence interval (CI). The ICC values were interpreted as follows: 0.00–0.20=poor agreement, 0.21–0.40=fair agreement, 0.41–0.60=moderate agreement, 0.61–0.80=substantial agreement, 0.80–1.00=excellent agreement.\(^{19}\) Spearman’s correlation coefficient was used to assess the relationship between age and the EBOS mean total score, and non-parametric Kruskal-Wallis ANOVA was used to assess the relationship between EB types/subtypes and the EBOS median total score. P-values of less than .05 were considered significant. Statistical analyses were performed using the SPSS software (SPSS for Windows, version 17.0; SPSS Inc, Chicago, IL – USA.)
RESULTS

Patients’ characteristics

During the study period ninety-two patients (48 [52.2%] females and 44 [47.8%] males) with a mean age of 15.4 years (range: 2 months–63 years) (SD: 14.1; 95% CI: 12.4 – 18.3) with different EB types/subtypes (Table 1) were tested with the EBOS.

Activity and Damage Score Distribution

The EBOS mean total score ± standard deviation (SD) was 12.9 ± 10.9 (range 0–33.5), while the mean total score ± SD for disease activity and structural damage was 10.1 ± 8.5 (range: 0–27.5) and 2.8 ± 2.6 (range: 0–6), respectively.

The range of EBOS mean score was 0-33.5 out of a possible 60 maximum score, with a median of 13.00 and an IQR of 20.50 (Table 2; Figure 2), whereas in the sub-sample (N=34) the range of mean score on time 1 (test) was 0 – 24.5, with a median of 13.00 and an IQR of 18.50 and on time 2 (re-test) 0 – 26.5 with a median of 11.50 and an IQR of 16.50 (Table 2; Figure 3). The highest EBOS score by site was measured for tongue, while the lowest for upper and lower fornices, either on total sample (Table 3) or sub-sample (Table 4).

Inter- and Intra-observer reliability

The inter- and intra-observer reliability for total score on all EB patients were excellent (ICC:0.94; 95%CI: 0.90–0.96 and ICC:0.90; 95%CI: 0.84–0.94, respectively). Even a comparison between the mean total score of RDEB patients versus all other EB patients showed an excellent agreement, unlike the total score on all other EB patients in the inter-rater assessment (ICC:0.58; 95%CI: 0.30–0.77) (Table 5).
Analyzing disease severity, either in the inter- or intra-rater assessment, lower lip reached the highest agreement (ICC:0.89; 95%CI: 0.84–0.92; ICC:0.91; 95%CI: 0.83–0.95, respectively), followed by tongue (ICC:0.87; 95%CI: 0.81–0.91) in the inter-rater, and by hard palate (ICC:0.87; 95%CI: 0.76–0.93) in the intra-rater (Table 3 and 4). As far as structural damage is concerned, microstomia and ankyloglossia either in the inter-rater (ICC: 0.93; 95%CI: 0.90–0.97; ICC: 0.89; 95%CI: 0.84–0.93, respectively), or in the intra-rater assessment (ICC: 0.94; 95%CI: 0.89–0.97; ICC: 0.94; 95%CI: 0.88–0.97, respectively) reached the highest agreement (Table 3 and 4).

**Correlation of EBOS score with EB types/subtypes and age.**

Lastly, a significant correlation was seen between EB types/subtypes and the EBOS median score (K-W ANOVA=71.626; \( P < .001 \)) (Table 2), but not between age and the EBOS mean total score. Indeed, there was a decline of score in EBS and JEB patients with age (EBS: \( \rho = -0.52, \ p = .107 \); JEB: \( \rho = -0.80, \ p = .13 \) ) and an increase in the rest of EB patients (DDEB: \( \rho = 0.20, \ p = .42 \); RDEB–O: \( \rho = 0.20, \ p = .48 \); RDEB-sev gen: \( \rho = 0.03, \ p = .85 \)), but none of them were statistically significant.
DISCUSSION

Since its first description in 1879,20 the oropharyngeal involvement in EB patients has always been one of the major concerns for every clinician, and they still represent one of the most important and challenging clinical manifestations, either in terms of evaluation, prognosis, and treatment. If, on the one hand, some medical/surgical treatments21-29 have been attempted in order to improve EB patients’ oral conditions, on the other hand, evaluation and prognosis of oropharyngeal component in EB still remain an enigma.

All EB types/subtypes may virtually experience all four clinical signs evaluated in this study, although with different frequency and extension: some had no or a very few and mild lesions and some others had many severe and disfiguring ones. Therefore, we thought that it was important to have an instrument capable to distinguish between active lesions and damage, in order to improve the global assessment of oropharyngeal severity either in hospital setting or in clinical trial of some potential medications. To the best of our knowledge, there is no published data on the grading of the individual clinical signs in each single oropharyngeal site in EB.

Our results on 92 EB patients showed that the EBOS has an excellent inter- and intra-observer reliability on both total and partial score of disease activity and structural damage score, with a median score changing significantly upon different EB type/subtypes (p < .001) (Table 2), unlike age which showed no correlation with the EBOS mean total score.

Considering also that we expected that RDEB patients were more severe, in order to reduce the bias of sample dilution, we divided all 92 patients in 2 sub-groups: all RDEB patients versus all other EB type/subtype patients. Even in this case the intra-rater reliability was excellent in both groups, while the inter-rater reliability was excellent for RDEB patients and moderate for all other EB patients (Table 5). It is likely that it is easier for 2 different physicians to detect gross and more widespread lesions, as
usually seen in RDEB patients, rather than subtle and sporadic lesions, as usually seen in the rest of EB patients. Interestingly, RDEB patients reached the highest EBOS mean total score versus all other EB types/subtypes (Figure 2) as supposed, in line with previous reports.\textsuperscript{6-10,18,21} Therefore, this result would support the EBOS validity.

The inter-rater reliability demonstrated to be poor-fair for upper and lower fornices (ICC: 0.05 and 0.10, respectively), floor of the mouth (ICC: 0.21) and oropharynx (ICC: 0.34) (Table 3), but was substantial-excellent either for disease activity or structural damage in the intra-rater assessment. Unfortunately, two sites (soft palate and oropharynx) and one permanent damage (enamel hypoplasia) were unable to be evaluated, since date were absent for both scorers or present just for one. All these disagreements between the 2 scorers might reflect their different capability to detect and classify a lesion. For instance, the different use of a light might have led to an under or over-estimation of some clinical signs, such as erythema, whose detection largely relies upon light direction, mostly in the most posterior oropharyngeal sites, like soft palate and oropharynx. Lastly, the EBOS was a very quick and easy to use tool, as time for record ranged from 1 to 5 minutes.

Two parameters were not included in the EBOS: first, the presence of oropharyngeal malignancy, because we thought that this needed a separate evaluation, considering how severe and worse in terms of prognosis could be. Also, to the best of our knowledge, there are only two reports of EB and oropharyngeal cancer\textsuperscript{30-31} and, therefore, this would have very likely resulted in adding an unvalidated parameter. If this means that EB patients very rarely develop oropharyngeal cancer or simply such lesions are undetected or underestimated remains unknown. In our study, none of 92 patients presented with any oropharyngeal malignancy and/or potentially malignant disorders.

Second, the presence of dental caries and/or periodontal disease (with subsequent dental loss) was not included, as not directly related to the disease. Indeed, although the percentage of dental caries seemed to be higher in EB patients versus a control group, it was not related to the disease.\textsuperscript{32} However,
it would be necessary to compare EB patients with a control group selected from different social background to really ascertain the incidence in both groups.

Lastly, we decided to not evaluate associated symptoms, such as pain and/or discomfort, and daily activities impairment, such as ability to eat, drink, or speak in this score, because these were considered subjective parameters to be evaluated in a study on the quality of life and quality of oral health.

This study has several limitations. The sample size was small and the number of EB subtypes was limited. It would be preferable to test this score in different hospital settings on a larger sample and higher variety of different EB subtypes. Another limitation could be the evaluation of inter- and intra-observer reliability by only two physicians. However, this was due to on purpose in order to not cause an excessive pain and unpleasant burden with multiple visits by more than two physicians. In addition, the intra-observer reliability was only tested on a part of the total sample.

Despite all these limitations, the EBOS may offer many important future perspectives. First, a validated scoring system capable of objectively evaluating the oropharyngeal disease severity might be useful even longitudinally to better understand the disease progression, thereby representing a valid prognostic tool. Indeed, the EBOS might be used to follow EB patients from their birth and assess how oropharyngeal cavity is going to be affected over time in order to possibly prevent any structural damage and ameliorate disease activity. Second, it appears that the severity of oropharyngeal lesions increase with the severity of cutaneous lesions\(^3\) and, therefore, it would be interesting to correlate the EBOS with other cutaneous score already present in the literature. Third, it would be interesting to correlate the EBOS with the type of mutations, in order to see whether or not there is an oropharyngeal phenotypic-genotypic correlation. Fourth, it would be useful to investigate any clinically meaningful change of the EBOS in response to medical and/or surgical oropharyngeal care interventions, particularly in clinical trials.
In conclusions, in this preliminary study the EBOS turned out to be a clinically valid and reliable tool to assess the oropharyngeal severity in EB patients. However, further investigations and refinements by other groups worldwide are strongly encouraged to better confirm our results.

ACKNOWLEDGMENTS

We are very grateful to all EB children and adults for their courage, strength, and patience. We would like to thank Dr. Phuong Khuu, Dr. Grainne O’Regan, all dermatology residents, Lisa Taylor, Lorraine Spaulding, Robin Newman, Gretchen Flanagan, Abigaille Augustin, and all EB clinic staff of the EB Clinic at Lucile Packard Children’s Hospital and the Bullous Disease Clinic at Stanford. We also would like to thank all Debra Mexico staff, specifically Lic. Erika Salas, Myrna Lozano, Juan Salas, Dr. Erika Garza and Lic. Peggy Perez.
REFERENCES


6 Wright JT. Oral manifestations in the epidermolysis bullosa spectrum. *Dermatol Clin* 2010; 28: 159-64.


11 Wright JT, Johnson LB, Fine JD. Development defects of enamel in humans with hereditary


Figure Legend.

**Figure 1.** Epidermolysis Bullosa Oropharyngeal Severity score (EBOS).

**Figure 2.** Box plot analysis of EBOS total score distribution per EB type/subtype calculated on full sample of 92 patients. Lines in box denote medians, bar include at most 1.5 of interquartile distance, difference between first and third quartile of score. Circle indicate value of the extreme outlier.

**Figure 3.** Box plot analysis of EBOS total score distribution calculated on a subsample of 34 patients at time 1 (test) and time 2 (re-test) per EB type/subtype. Lines in box denote medians, bar include at most 1.5 of interquartile distance, difference between first and third quartile of score. Circle indicate value of the extreme outlier.
<table>
<thead>
<tr>
<th>Patient N°</th>
<th>Date of birth (year)</th>
<th>Sex</th>
<th>Initials</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Name of Physician</th>
<th>Date</th>
</tr>
</thead>
</table>

### EXTENT OF DISEASE ACTIVITY

<table>
<thead>
<tr>
<th>Sites involved</th>
<th>Erythema</th>
<th>Atrophy</th>
<th>Blister</th>
<th>Erosion/Ulcere</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Upper lip</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Lower lip</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Upper fornix</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Lower fornix</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Upper gingiva</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. Lower gingiva</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. Hard palate</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. Soft palate</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9. Left cheek</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10. Right cheek</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11. Tongue</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12. Floor of the mouth</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13. Oropharynx</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Partial score: 13/13/13/13/52

### EXTENT OF STRUCTURAL DAMAGE

<table>
<thead>
<tr>
<th>Lesions</th>
<th>Microstomia</th>
<th>Ankyloglossia</th>
<th>Intramuscular muscle damage</th>
<th>Enamel hypoplasia</th>
</tr>
</thead>
<tbody>
<tr>
<td>Absent</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Present</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Partial score: 2/2/2/2/8

**TOTAL SEVERITY SCORE**: 60

Mark (√) each applicable blank. For Disease activity score, each check mark = 1 point. For Structural Damage, Absent = 0 point and Present = 2 points. Tally total for final score.