Acta Odontologica Scandinavica. 2015; 73: 156–160

informa
healthcare

SHORT COMMUNICATION

Evaluation of internal consistency of the epidermolysis bullosa oropharyngeal severity score (EBOS)

GIULIO FORTUNA[1,2], MASSIMO ARIA[3], RODRIGO CEPEDA-VALDES[2] & JULIO CESAR SALAS-ALANÍS[2,4,5]

[1]*Department of Oral Medicine, Carolinas Medical Center, Charlotte, NC, USA,* [2]*D.eb.RA. Mexico Foundation, Monterrey, Nuevo Leon, Mexico,* [3]*Department of Economics and Statistics, Federico II University of Naples, Naples, Italy,* [4]*Dermatology Service, Instituto Tecnológico y de Estudios Superiores de Monterrey, Monterrey, NL, México, and* [5]*Department of Basic Science, Universidad de Monterrey, San Pedro Garza García, NL, México*

Abstract

Objective. To evaluate the internal consistency of the epidermolysis bullosa oropharyngeal severity score (EBOS). ***Materials and methods.*** Data from 92 patients of varying EB types/sub-types already described in a previous multi-center study were re-analyzed via the coefficient Cronbach's α (CR-α). Additionally, the corrected item total correlation between each item and the items' overall score with Pearson's product-moment correlation (ρ) was calculated. ***Results.*** The alpha coefficient for the mean total score of 17 items is 0.941. The inter-observer reliability for disease severity score was excellent for oral medicine specialist ($\alpha = 0.924$) and dermatologist ($\alpha = 0.916$) and the intra-observer reliability was good at Time 1 ($\alpha = 0.895$) and Time 2 ($\alpha = 0.897$). The analysis of CR-α per single item revealed that alpha was greater than 0.904 for disease activity and 0.743 for structural damage, after the elimination of four items for oral medicine specialist and greater than 0.898 for disease activity and 0.769 for structural damage after the elimination of five items for dermatologist. Similarly the analysis of the corrected items-EBOS correlation showed that the same items do not correlate very well ($\rho < 0.4$) with the overall EBOS. ***Conclusions.*** The EBOS turned out to have a strong and reliable internal consistency, as the majority of the EBOS' items were consistent with each other.

Key Words: *EB, consistency, score, validity, reliability*

Introduction

We have recently published a study on a new instrument, the epidermolysis bullosa oropharyngeal severity score (EBOS), capable of quantifying as much as objectively possible the oral-pharyngeal involvement in patients with epidermolysis bullosa [1]. Our investigation was aimed at assessing the EBOS' inter- and intra-observer reliability, calculated by the intra-class correlation coefficient (ICC) [2], which is one of the methods for evaluating the degree of agreement or consensus made by the same or different observers [3,4].

The results of our previous study [1] indicated that both inter- and intra-observer reliability for total score on all EB patients were considered excellent. Similarly, even when analyzing each single parameter of the disease activity and structural damage, a substantial–excellent correlation was found in the inter-observer (except for four out of 13 sites) and intra-observer reliability [1].

Here, we have re-performed the statistical analysis of the same rough data in order to evaluate the EBOS' internal consistency, that is a measure based on the correlations between different items on the same test. So, we measured the EBOS' consistency in relation to the hypothesis of unidimensionality, i.e. all items measure the same phenomenon under examination (the oral-pharyngeal severity of the disease). Thus, we verified whether or not all items were proposed to measure the same general construct and were, therefore, consistent with each one other. The EBOS' internal consistency was evaluated by one of the most widely used methods to ascertain the degree

Correspondence: Giulio Fortuna, DMD PhD, Chief Resident, Department of Oral Medicine, Carolinas Medical Center, 1000 Blythe BLVD, PO Box 32861, Charlotte, NC 28232-2861, USA. Tel: +1 980 318.2006. Fax: +1 704 355.5301. E-mail: giulio.fortuna@carolinashealthcare.org; giulio.fortuna@gmail.com

(Received 14 December 2013; accepted 22 May 2014)

ISSN 0001-6357 print/ISSN 1502-3850 online
DOI: 10.3109/00016357.2014.931460

of correlation between variables, that is, the coefficient Cronbach's α (CR-α) [4,5].

This new analysis was aimed at determining which of the items were useful for the EBOS' internal consistency, that is, whether or not all the EBOS items were really necessary in order to have a reliable severity scale. The analysis with the Cronbach's alpha (and the item-scale correlation) should be viewed as a completion of the assessment of the validity of the EBOS, in addition to the analysis of the inter- and intra-reliability already performed [1], as we thought that it was mandatory to not confuse the two concepts (reliability and internal consistency), but rather investigate both of them and confirm the EBOS validity, before using it in clinical trials.

Materials and methods

The methodology of collection of the rough data from 92 patients of varying EB types/sub-types has already been described in our previous study, for which we obtained ethical approval from Instituto Tecnológico (Mexico) and Stanford University (USA) [1]. Essentially, each of 92 EB patients received a complete clinical intra-oral examination by two examiners (dermatologist and oral medicine specialist), who scored them independently, not simultaneously. These two different oral healthcare providers made the assessment of oral severity, composed by 'disease activity' and 'structural damage' in 13 oral anatomical sites [1], based on their own personal medical/dental training/background. No pre-training or pre-calibrating test was done in order to not introduce any bias. Such rough data, previously analyzed for calculating the inter- and intra-observer reliability with the ICC, have been now re-analyzed using CR-α for assessing the EBOS' internal consistency.

We have also calculated the corrected items-EBOS correlation, viz. the correlation between each item and a composite score of all the other remaining items minus the item of interest. The CR-α values were interpreted as follows: $\alpha \geq 0.9$ = excellent; $0.8 \leq \alpha < 0.9$ = good; $0.7 \leq \alpha < 0.8$ = acceptable; $0.6 \leq \alpha < 0.7$ = questionable; $0.4 \leq \alpha < 0.6$ = poor; and $\alpha < 0.4$ = unacceptable. Internal consistency was calculated for each dimension of disease activity and structural damage, which was expected to have an alpha ≥0.7, the minimum recommended value to reflect a relationship between items within a score [6]. If a dimension's alpha increased after an item had been eliminated, this item was considered of questionable/poor value for the scale. The corrected item total correlation between each item and the items' overall score was calculated with Pearson's product-moment correlation (ρ). A correlation of $\rho \geq 0.4$, after the item was excluded, was considered acceptable [6]. Therefore, an item was considered to be not consistent with the overall scale, if both conditions were fulfilled: increased CR-α and ρ lower than 0.4.

Results

Our results showed that the internal consistency was very high for the EBOS. Indeed, the inter-observer reliability for disease severity score was excellent for oral medicine specialist ($\alpha = 0.924$) and dermatologist ($\alpha = 0.916$), as well as the disease severity mean score ($\alpha = 0.941$), whereas the intra-observer reliability for disease severity score was good at Time 1 ($\alpha = 0.895$) and Time 2 ($\alpha = 0.897$) (Table I), confirming the results of our previous work [1].

However, the analysis of CR-α per single item revealed that alpha increased after the elimination of four items: upper fornix ($\alpha = 0.909$), lower fornix ($\alpha = 0.908$), floor of the mouth ($\alpha = 0.911$) and enamel hypoplasia ($\alpha = 0.909$), for oral medicine specialist, and of five items: upper fornix ($\alpha = 0.906$), lower fornix ($\alpha = 0.907$), floor of the mouth ($\alpha = 0.906$), oropharynx ($\alpha = 0.903$) and enamel hypoplasia ($\alpha = 0.909$) for dermatologist. Additionally, the analysis of the corrected items-EBOS correlation showed that four specific items (upper fornix, lower fornix, floor of the mouth and enamel hypoplasia) do not correlate very well ($\rho < 0.4$) with the overall EBOS, either for oral medicine specialist or dermatologist (Table II).

Such low correlation is particularly evident on the floor of the mouth, where the scatter plot analysis

Table I. Internal consistency of the EBOS using Cronbach-α. Inter-observer reliability of disease activity score calculated on a full sample size ($n = 92$) and intra-observer reliability on a sub-sample of patients ($n = 34$) calculated on the first (Time 1) and second round (Time 2).

	Inter-observer using Cronbach-α			Intra-observer using Cronbach-α		
EBOS	Oral medicine specialist	Dermatologist	Mean score	Time 1*	Time 2*	Number of items evaluated
Disease activity	0.904	0.898	0.912	0.890	0.889	13
Structural damage	0.743	0.769	0.768	0.772	0.728	4
Disease severity (total score)	0.924	0.916	0.941	0.895	0.897	17

*Time 1 and Time 2 indicate the mean score between an oral medicine specialist and dermatologists.

Table II. Evaluation of internal consistency per single item of the epidermolysis bullosa oropharyngeal severity score (EBOS).

	Cronbach-α (if an item is deleted)		Corrected Items – EBOS correlation	
EBOS	Oral medicine specialist	Dermatologist	Oral medicine specialist	Dermatologist
Disease activity				
Upper Lip	0.889	0.893	0.777	0.691
Lower Lip	0.890	0.891	0.756	0.734
Upper fornix	0.909*	0.906*	0.218**	0.271**
Lower fornix	0.908*	0.907*	0.230**	0.225**
Upper Gingiva	0.893	0.895	0.717	0.689
Lower Gingiva	0.892	0.895	0.748	0.651
Right Cheek mucosa	0.887	0.889	0.815	0.774
Left Check mucosa	0.889	0.888	0.781	0.786
Hard Palate	0.887	0.894	0.801	0.673
Soft Palate	0.901	0.898	0.516	0.609
Tongue	0.890	0.892	0.807	0.796
Floor of the mouth	0.911*	0.906*	0.026**	0.311**
Oropharynx	0.901	0.903*	0.538	0.431
Structural damage				
Microsotomia	0.472	0.532	0.838	0.850
Ankyloglossia	0.503	0.526	0.799	0.857
Other scars	0.591	0.638	0.681	0.699
Enamel HYPO	0.909*	0.910*	0.114**	0.180**

*Item deletion increases the alpha measure compared to the global scale value.
**A correlation (ρ) $\leq$0.4 was considered weak.

did not reveal any positive correlation between the mean disease activity score of the item (floor of the mouth) and the disease severity score (Figure 1).

Discussion

The alpha coefficient for the mean total score of 17 items is 0.941, suggesting that the items have a

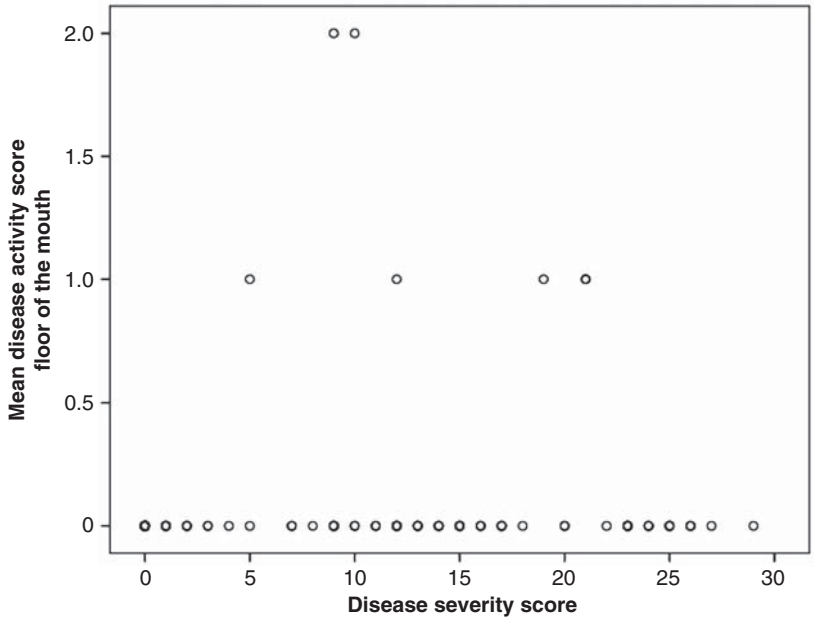

Figure 1. Scatter-plot analysis of disease severity score and mean disease activity score between oral medicine specialist and dermatologist measured for the floor of the mouth on total sample size.

very high internal consistency for both physicians (Table I). Similarly, excellent–good consistency was obtained for total score of either inter- or intra-observer reliability. Thus, it appears that all items are measuring a common construct.

Table II shows the values that CR-α would be if a particular item were deleted from the scale. If we consider the results of the analysis of the corrected item–EBOS correlation, we may notice that either for dermatologist or oral medicine specialist only four items (upper fornix, lower fornix, floor of the mouth and enamel hypoplasia) have a weak correlation (Table II). This is in line with our previous study in which we demonstrated a low ICC for the same items, except for enamel hypoplasia. However, despite their different background, dermatologists and oral medicine specialists seem to show the same accuracy in the overall evaluation of the oral-pharyngeal disease severity in EB patients.

Therefore, only those four items are apparently not measuring the same construct as the rest of the EBOS' items do and should be removed, as we arbitrarily fixed the acceptable ρ value equal to or higher than 0.4. However, this is not a universal value, as in many other studies it was fixed at 0.2 [7] or 0.3 [8]. So, before eliminating such items (as in the case of the first three EBOS' items), we may then claim that those items with corrected item-total correlations between 0.2–0.4 require more careful investigations with more multi-center prospective studies to better ascertain whether or not such items really represent a true measure of the scale.

The low correlation of the first three items (upper fornix, lower fornix and floor of the mouth) in the disease activity might be explained by the fact that about half of the EB population under examination was affected by a recessive form, in which diffuse a oral-pharyngeal scarring phenotype is quite frequent [9]. Certainly, this is the result of previous disease activity: EB patients experienced blisters and erosions first, then developed scar tissues, resulting in vestibule obliteration and ankyloglossia [9], with subsequent 'stabilization' of the disease and less symptoms just in those areas. As the disease *per se* leads to the disappearance of upper and lower gingival fornices and floor of the mouth over time, all these three sites become impossible to be scored, which is why we probably had a low correlation, making those items apparently not consistent with the unidimensionality of our scale. Conversely, the item oropharynx, despite an increased alpha value just for the dermatologist, was considered consistent with the EBOS, as its correlation (ρ) was higher than 0.4.

It also appears that the more severe the oral involvement, the more difficult the evaluation of these three sites and the more difficult to establish any positive correlation between disease severity and the mean disease activity of each item. Indeed, for instance, the majority of EB patients have a score of 0 on the floor of the mouth, no matter how severe the oral-pharyngeal disease was: the cloud of point in our analysis (Figure 1) did not show any clear direction, which means no correlation.

Therefore, not only a score-dependent component (different background), but probably also a disease-dependent component (scarring phenotype), may affect the reliability and validity of certain items of the EBOS.

Only those items with corrected item-total correlations of less than 0.2 should be removed from the instrument, for they would not be measuring the same construct with other items or discriminating subjects' characteristics that are supposed to be measured. This is precisely the case of enamel hypoplasia (Table II).

Apparently, the item enamel hypoplasia does not really measure the same phenomenon (disease severity), as the other items do, and should be either evaluated separately in another scale or considered as a different measure. It seems that this item does not explain the unidimensionality of the phenomenon under examination, probably because enamel hypoplasia is not correlated to a previous disease activity, but directly to the altered genes, which cause structural damage of the enamel since the beginning [10]. We may hypothesize that this might be one of the reasons that enamel hypoplasia reached such a very low correlation value, despite its excellent inter- and intra-observer reliability measured by ICC as previously demonstrated [1], since CR-α and ICC measure two different aspects of a scale. Additionally, we need to consider that enamel hypoplasia is the only item which refers to an involvement of the hard tissues of the oral cavity, whereas the rest of the EBOS' items (either those related to disease activity or structural damage) refer to a soft tissue involvement. On the other hand, enamel hypoplasia was found and evaluated in only a small number of EB patients (four out of 92 EB patients). The small number of cases in some lesions truly represents a limitation of our analysis, which might influence the results of the ICC and alpha.

Despite this, our results showed that the scale (EBOS) is very well built, looking at the alpha correlation and the already evaluated stability of the inter- and intra-rater reliability. Additionally, it should also be noted that, while a low value for CR-α indicates poor internal consistency of the items in the scale, this does not necessarily imply that the scale is not unidimensional.

We may conclude that (i) the EBOS has a strong and reliable internal consistency, as the majority of the EBOS' items are consistent with each other; (ii) gingival fornices and floor of the mouth should be better re-analyzed in a more homogenous cohort of EB patients, with more numerous and diverse scorers, to better confirm whether or not they should be definitively removed from the EBOS scale; and (iii)

enamel hypoplasia should be evaluated in a larger EB population presenting with this type of structural damage, mainly Junctional EB patients, before deciding definitely whether enamel hypoplasia is a measure of the oral-pharyngeal disease severity or simply represents another aspect/dimension of the complexity of oral-pharyngeal issues in EB patients.

***Declaration of interest*:** The authors report no conflicts of interest. Contents of the manuscript have not been previously published or presented in a congress and are not currently submitted elsewhere.

References

[1] Fortuna G, Chainani-Wu N, Lozada-Nur F, Aria M, Cepeda-Valdes R, Pollio A, et al. Epidermolysis Bullosa Oropharyngeal Severity (EBOS) score: a multicenter development and reliability assessment. J Am Acad Dermatol 2013;68: 83–92.

[2] Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977;33:159–74.

[3] Muller R, Buttner P. A critical discussion of intraclass correlation coefficients. Stat Med 1994;13:2465–76.

[4] Cronbach LJ. Coefficient alpha and the internal structure of tests. Psychometrika 1951;16:297–334.

[5] Cronbach LJ, Shavelson RJ. My current thoughts on coefficient alpha and successor procedures. Educ Psychol Meas 2004;64:391–418.

[6] Hagberg K, Branemark R, Hagg O. Questionnaire for persons with a Transfemoral Amputation (Q-TFA): initial validity and reliability of a new outcome measure. J Rehabil Res Dev 2004; 41:695–706.

[7] Erhart M, Ellert U, Kurth BM, Ravens-Sieberer U. Measuring adolescents' HRQoL via self reports and parent proxy reports: an evaluation of the psychometric properties of both versions of the KINDL-R instrument. Health Qual Life Outcomes 2009;7:77.

[8] McMillan CV, Honeyford RJ, Datta J, Madge NJ, Bradley C. The development of a new measure of quality of life for young people with diabetes mellitus: the ADDQoL-Teen. Health Qual Life Outcomes 2004;2:61.

[9] Wright JT. Oral manifestations in the epidermolysis bullosa spectrum. Dermatol Clin 2010;28:159–64.

[10] Wright JT, Hall KI, Deaton TG, Fine JD. Structural and compositional alteration of tooth enamel in hereditary epidermolysis bullosa. Connect Tissue Res 1996;34:271–9.