A novel comprehensive scoring system for oral lichen planus: A validity, diagnostic accuracy, and clinical sensitivity study

Hossam H. Elsabagh, BDS, MSc,[a] Yasmine Y. Gaweesh, BDS, MSc, PhD,[b] Jaylane K. Ghonima, BDS, MSc,[a] and Marwa Gebril, BDS, MSc[a]

Objectives. The objective of this study was to establish and validate a new scoring system for oral lichen planus (OLP).

Study Design. Forty patients with erosive OLP were scored using a new proposed scoring system and the one suggested by Thongprasom. Statistical analysis was used to calculate inter- and intraexaminer reliability, validity, diagnostic accuracy, and clinical sensitivity for both scores.

Results. Concurrent validity showed a significant strong correlation between pain scale and the new score (r_s = 0.846) versus Thongprasom's score (r_s = 0.665). Interexaminer agreement showed a statistically significant agreement with the biopsy results (kappa = 0.74) for the new score, whereas no agreement was evident with Thongprasom's score (kappa = 0.03163). The diagnostic accuracy of the new score was area under the curve = 0.839, whereas the diagnostic accuracy of Thongprasom's score was area under the curve = 0.667. Significant differences in the scores were detected while comparing the clinical sensitivity of both systems.

Conclusions. The new comprehensive scoring system could be a valid, reproducible, and sensitive tool to accurately assess the severity of OLP. In addition, the proposed scoring system is easily taught, is relatively faster to master, and does not require complicated calculations. (Oral Surg Oral Med Oral Pathol Oral Radiol 2021;131:304−311)

Oral lichen planus (OLP), one of the most common diseases affecting the oral tissues, remains the focus of extensive research. With an approximate global prevalence of up to 2%, this chronic inflammatory disease is a cause of marked morbidity for many patients. In addition, it is documented to be a potentially malignant disorder.[1]

Although the pathogenesis of OLP has been thoroughly investigated, there is no consensus on the exact etiology of OLP. The present data suggest that OLP is a T-cell-mediated autoimmune disease in which cytotoxic $CD8^+$ T cells elicit apoptosis of basal keratinocytes, disturb basement membrane integrity, and trigger chronic inflammation.[2] According to the degree of this persistent inflammation, OLP typically exhibits identifiable clinical characteristics with a distinctive distribution. It could manifest in 3 different patterns that occur individually or in combination: erythematous (atrophic), erosive (ulcerated, bullous), and white keratotic lesions. The latter represents the recognizable clinical feature of OLP that may arise in the form of white papular, reticular, or plaque−like patterns that usually involve both buccal mucosae. This keratotic form is usually asymptomatic and can be considered the ultimate form reached after treatment.[3]

Erythematous, eroded, and ulcerative forms of OLP are commonly accompanied by white striae. These striae help to clinically differentiate OLP from other vesiculo−erosive diseases that are featured with isolated areas of erythema and/or erosions. Destructive patterns of OLP (atrophy, erosions, and ulcerations) usually cause varying degrees of pain and discomfort, which act as the main motivating force for seeking treatment.[4,5]

A wide variety of treatment options have been investigated, all aiming at pain relief and reduction of lesion size. Therefore, a variety of scoring systems have been proposed to evaluate disease severity and monitor the response to such treatments, testing the effectiveness of these drugs within and among patients. To date, more than 22 specific scoring systems have been reported, in addition to a number of nonspecific oral disease scoring systems that could also be used for monitoring OLP.[6,7]

Although many of these scoring systems have already been used over the past 3 decades, only a few have been sufficiently validated.[8] None has been investigated for its diagnostic accuracy, sensitivity, or specificity. In addition, some of these systems require a considerable degree of experience to ensure the reliability and reproducibility of the score. Furthermore, there is a compelling need to improve such scales to enable more precise and easy judgment of disease severity and consequent proper assessment of patients' response to therapy. Moreover, standardization of the scoring system among various clin-

Statement of Clinical Relevance

The new proposed scoring system could be used as a comprehensive scoring tool for routinely monitoring severity of oral lichen planus in clinical practice as well as comparing different treatment modalities reported in randomized clinical trials.

[a]Assistant Lecturer, Oral Medicine and Periodontology, Alexandria University, Egypt.
[b]Lecturer, Oral Medicine and Periodontology, Lecturer, Alexandria University, Egypt.
Received for publication Aug 16, 2020; returned for revision Dec 14, 2020; accepted for publication Dec 17, 2020.

2212-4403/$-see front matter
https://doi.org/10.1016/j.oooo.2020.12.016

ical trials (which test new treatment modalities) would allow a more thorough comparison of their results.

Therefore, the objective of this study was to propose a new comprehensive scoring system of OLP to fill in the gaps of the previous scores and minimize the error margin in describing the actual disease severity.

MATERIALS AND METHODS

Participants, materials, and study design

The present study was conducted following the principles of the modified Helsinki's code for human clinical studies (2013), the guidelines of STARD 2015 for reporting diagnostic accuracy studies, and the guidelines of the Research Ethics Committee of the Faculty of Dentistry, Alexandria University, Egypt (Institutional Review Board No. 00010556-IORG0008839). Between January 2019 and February 2020, 40 consecutive adult patients suffering from biopsy-proven OLP were included in the study. All patients provided written informed consent, and all were recruited from the outpatient clinic of the Oral Medicine, Periodontology, Diagnosis, and Radiology Department, Faculty of Dentistry, Alexandria University, Egypt. Patients were excluded if they had desquamative gingivitis caused by a vesiculobullous disease other than OLP.

When proven eligible, patients were scored using 2 different scoring systems, as explained in detail in the following section: the new proposed score and the score suggested by Thongprasom et al.,[9] which is the most commonly used in the literature.

Thongprasom scoring system

Scoring using Thongprasom's system[9] is as follows: 0 = normal mucosa; 1 = a lesion having only white striae; 2 = a lesion of white striae and atrophic areas <1 cm^2; 3 = a lesion of white striae and atrophic areas >1 cm^2; 4 = a lesion of white striae and erosive areas <1 cm^2; and 5 = a lesion of white striae with erosive areas >1 cm^2.

New scoring system (Elsabagh et al.'s scoring system for OLP)

The newly proposed scoring system comprises 4 separate categories that cover all oral criteria of the disease. Each category is given a subscore, with all subscores summed to obtain the final score of the patient. These categories are as follows:

1. Objective mucosal lesion nature (no lesion = 0, white keratotic lesion = 1, atrophy/erosion intermixed or not with white lesion = 2, ulceration intermixed or not with white lesion = 3).
2. Subjective pain score (no pain = 0, mild pain = 1, moderate pain = 2, severe pain = 3).
3. Number of surfaces affected in the oral cavity other than the gingiva (only 1 surface affected or buccal mucosae bilaterally = 0, more than 1 surface affected or more than both buccal mucosae = 1).
4. Gingival involvement as desquamative gingivitis (no gingival involvement = 0, narrow band [1 mm] of gingival involvement or wide band in less than 6 teeth involved = 1, wide band [>1 mm] of gingival involvement in more than 6 teeth involved = 2).

Thus, in the new scoring system, final scores (i.e., the sum of all subscores) ranges from 0 to 9, with the worst disease severity as 9, whereas a score of 0 represents complete resolution of the disease. Figure 1 illustrates this new proposed scoring system.

The second item in the new score (subjective pain score) was recorded using a numeric rating scale (NRS).[10] Patients were asked to verbally assign a numerical score on the scale to rate their pain intensity, and the number was recorded and then categorized as 0 = no pain, 1 to 3 = mild pain, 4 to 7 = moderate pain, and 8 to 10 = severe pain.[11,12]

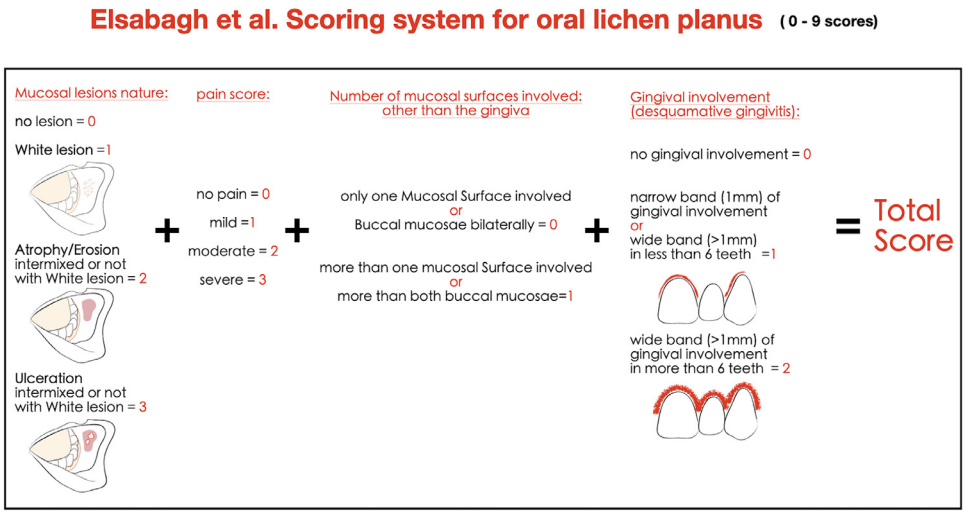

Fig. 1. The new proposed scoring system (Elsabagh et al.'s scoring system for oral lichen planus).

Outcome measures

Reliability. During the initial examination (baseline), the scoring of every patient was performed clinically, once according to Thongprasom's score and once using the new proposed score, by 3 calibrated investigators (H.E., Y.G., and J.G.) independently to assess interobserver reliability. Subsequently, a set of full-mouth photographs was taken from each patient including the upper and lower labial mucosae; right and left buccal mucosae; dorsal, ventral, and lateral surfaces of the tongue; floor of the mouth; hard and soft palate; teeth; and surrounding gingiva until the vestibules. The scoring process was repeated after 1 week to assess only intraobserver reliability.

Validity. Multiple statistical analyses were applied to the data collected to compare both scoring systems with regard to their concurrent validity, where NRS was correlated with both scores and to test the construct validity of the new scoring system.

Diagnostic accuracy and interexaminer agreement. Histopathologic examination of the confirmatory biopsies obtained from the most destructed sites of each patient was performed under a light microscope by a professional oral pathologist, who was blinded to clinical scores, to locate sites with hyperkeratosis, atrophy, erosion, or ulcer and confirm the nature of the disease. Atrophy was stated when up to 25% of the epithelium was lost, erosion was stated when more than 25% of epithelium was lost but basement membrane was still intact, and ulcer was stated when the whole thickness of epithelium was lost along with the basement membrane. Then, correlation of the biopsy results with both Thongprasom's score and objective mucosal lesion nature of the proposed scoring system (assessed blindly from histopathology results) was done to examine the agreement between both scoring systems and biopsy results. Furthermore, the diagnostic accuracy of scores was tested in terms of sensitivity and specificity.

Clinical sensitivity. After baseline examination and scoring, patients were prescribed a treatment regimen of topical corticosteroid (Kenacort−A−Orabase, triamcinolone acetonide 0.1% oral gel, DEVA Pharmaceuticals, Çerkezköy,Turkey) and topical antifungal agent (Miconaz oral gel, Amoun Pharmaceutical Co., El Obour City, Cairo, Egypt), and systemic corticosteroids were also prescribed for patients who were suffering from severe pain. Patients were recruited for follow-up 1 month and 3 months after starting medication. At each follow-up time point, patients were clinically reexamined and scored using the 2 scoring systems. Statistical analysis was used to compare the ability of both scoring systems to detect even minor clinical improvements.

Sample size calculation

A sample size of 40 patients achieves 90% power to detect a correlation between scores, assuming a correlation coefficient of 0.5. Two-sided correlation tests with a significance level of .05 were performed. The sample size was calculated using MedCalc software (version 12.4.0, Belgium).

Statistical analysis

A Kruskal-Wallis *H* test was applied to detect differences between the total score between the 3 raters. Inter− and intraobserver reliabilities were assessed. Assessment for the level of agreement in terms of the intraclass correlation coefficient (ICC) for ordinal or continuous measures followed well−established benchmark limits (Fleiss's and Altman's benchmark scales).

A Bland-Altman plot was used to assess the agreement between the new score and Thongprasom's score and to estimate the bias and limits of agreement.

We also tested construct validity, which is the extent to which a particular measure performs in accordance with theoretical expectations. In this study, the scores for symptoms of OLP could be expected to increase as the score increased for clinical signs. Spearman's correlation coefficient was estimated for every individual item of the new score measuring both symptoms and signs. The same test was used to test for the correlation between NRS and both scoring systems, and the kappa test for agreement was used to assess the agreement between biopsy results and both scores.

We used receiver operator characteristic curve analysis to assess the diagnostic accuracy of both scores, where keratosis/atrophy was considered low severity disease (negative outcome) and erosion/ulcer was considered high severity disease (positive outcome), with the calculation of sensitivity (true positive rate) and specificity (true negative rate). McNemar's test was used to compare the clinical sensitivity of both scores.

We performed statistical analysis using SPSS (version 20, IBM, Chicago, IL) and MedCalc. The significance level was set at $P < .05$.

RESULTS

Forty patients diagnosed with OLP were enrolled in this study. The mean (SD) age of the patients was 49.50 (7.31) years. No significant differences were observed among patients with regard to age or gender distribution.

Reliability

Interobserver and intraobserver reliability. Total scores of the new scoring system recorded by different observers showed no significant difference ($P = .882$). Using ICC, the new scoring system produced slightly higher interobserver reliability (0.97; 95% confidence

Table I. Interobserver reliability for both scoring systems and individual components of the new scoring system

Scoring system	*Range*	*Mean (SD)*	*Median (IQR)*	*ICC (95% CI)*	*P value*	*Overall benchmark value*
Lesion severity	1-3	2.30 (0.61)	2 (2-2)	0.89 (0.81-0.94)	<.0001*	Almost perfect
Pain scores	0-3	1.75 (0.81)	2 (1-1)	1.00	—	Perfect
Site score	0-1	0.87 (0.33)	1 (1-1)	0.95 (0.92-0.97)	<.0001*	Almost perfect
Gingival	0-1	0.08 (0.27)	0 (0-0)	1.00	—	Perfect
Total new score	2-8	4.93 (1.43)	5 (4-4)	0.97 (0.95-0.98)	<.0001*	Almost perfect
Thongprasom's score	1-5	3.78 (1.30)	4 (3-3)	0.93 (0.88-0.96)	<.0001*	Almost perfect

IQR, interquartile range; ICC, intraclass correlation coefficient; CI, confidence interval.
*Statistically different at $P \leq .05$.

interval [CI], 0.95-0.98) than Thongprasom's score (0.93; 95% CI, 0.88-0.96), as shown in Table I. Additionally, test-retest (intraobserver) reliability of the new scoring system and Thongprasom's score showed high reliability for both (0.98; 95% CI, 0.97-0.99; and 0.96; 95% CI, 0.93-0.98, respectively).

Difference and limits of agreement for the new score versus Thongprasom's score. Figure 2 shows Bland-Altman plots in which the difference between the 2 scoring systems is plotted vs the means for both. The mean difference between the 2 scores was −1.15, and the limits of agreement were (−2.96 to 0.66). Based on this small difference, it can be assumed that both scoring systems could be used interchangeably.

Validity

Concurrent validity between the new scoring system and Thongprasom's score in relation to pain score reported by NRS. As shown in Table II, the results showed a statistically significant stronger correlation between NRS and the new score (r_s = 0.846; $P <$ .0001) versus Thongprasom's score (r_s = 0.665). This might validate the concurrent use of the new score as representative of pain and discomfort.

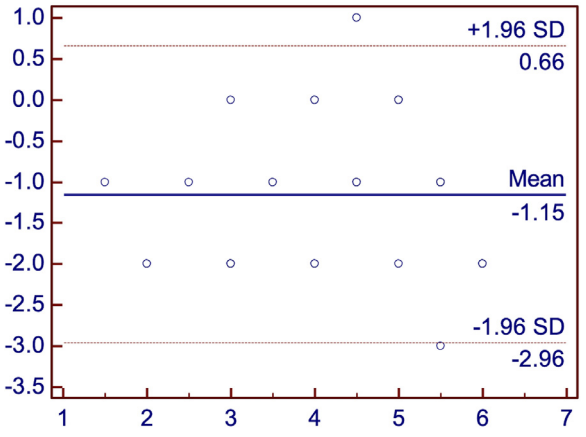

Fig. 2. Bland-Altman plot of the new score versus Thongprasom's score.

Table II. Concurrent validity between NRS and both the new scoring system and Thongprasom's score

		Correlation	
			NRS
Spearman's rho	NRS	r_s	1.000
		—	—
		N	40
	Thongprasom's score	r_s	0.665
		—	.000*
		N	40
	New score	r_s	0.846
		—	.000*
		N	40

NRS, numeric rating scale.
*Statistically different at $P \leq .05$.

Construct validity between the individual components comprising the new scoring system. As shown in Table III, construct validity testing showed not only that each item of the new score had a statistically significant correlation to the total score (strong correlation for lesion nature, pain, and number of sites and weak positive correlation for gingival involvement) but that there was a statistically significant correlation between the pain and both nature of lesion and gingival involvement, showing the coherence between the individual components of the new scoring system.

Interexaminer agreement and diagnostic accuracy

Interexaminer agreement (kappa) between the biopsy results (actual disease nature) and both Thongprasom's score and the new score (objective disease nature).
Interexaminer agreement between each of the scoring systems and the biopsy results was assessed. Unfortunately, 11 biopsy specimens, although diagnostic for OLP, were not indicative of the nature of the lesion and it was difficult to make an accurate determination on the intact thickness of the epithelium; thus, these 11 specimens were excluded.

As shown in the first part of Table IV, there was almost no agreement between Thongprasom's score

Table III. Construct validity between individual components of the new scoring system, along with total score

	r_s	*P value*
Lesion nature vs pain	0.66	<.0001*
Lesion nature vs total	0.83	<.0001*
Pain vs gingival	0.21	.019*
Pain vs total	0.89	<.0001*
Site vs total	0.24	.008*
Gingival vs total	0.33	.001*

*Statistically different at $P \leq .05$.

and the biopsy results (kappa = 0.03163, $P > .05$), with a total percentage agreement of 24.1% (7/29).

Furthermore, Thongprasom's score showed partial misperception in detection of both atrophy (3/7 detected) and erosion (3/8 detected), with complete failure to detect ulcerations (0/13).

On the contrary, the new score (objective disease nature), as shown in the second part of Table IV, showed statistically significant substantial agreement with the biopsy results (kappa = 0.74, $P < .05$), with a total percentage agreement of 86.2% (25/29).

Diagnostic accuracy of both Thongprasom's score and the new scoring system. The diagnostic accuracy of Thongprasom's score to discriminate between biopsy results was not statistically significant ($P = .192$), with an area under the curve (AUC) = 0.667, sensitivity of 80.95%, and specificity 50%. In contrast, the diagnostic accuracy of the new score was statistically significant ($P < .0001$) with an AUC = 0.839, sensitivity of 57.14%, and specificity 100%. Figure 3 shows the receiver operating characteristic curve for both scoring systems.

Clinical sensitivity

Using McNemar's test, there was statistically significant difference between scores from baseline to 1-month follow-up (Thongprasom's score = 73.3%; new score = 100%; ($P < .001$), though there was no statistically significant difference between scores from 1-month to 3-month follow-up (Thongprasom's score = 50%; new score = 66.7%; $P = .226$).

DISCUSSION

OLP has always been and will be the focus of significant research, because it is considered one of the most prevalent oral mucosal diseases. Moreover, no definite cure for the disease is available; instead, all therapy basically consists of treatment of symptoms. In a systematic review studying the different treatment modalities for OLP, Lodi et al.[13] pointed out that the lack of uniformity in the outcome measures of various studies was found to be a major obstacle in pooling and comparing the results of such studies, as highlighted in other studies.[14,15] The need for a standardized outcome measure in the form of a universal comprehensive scoring system has been emphasized in multiple systematic reviews and studies.[6,7,13,14]

Studies in the literature have proposed numerous scoring systems for OLP. Among these, the one suggested by Thongprasom et al.[9] was found to be the most commonly used in clinical trials.[16-19] Scrutinizing the various specific scoring systems, it could be concluded that Thongprasom's score, although deficient in multiple aspects, is typically preferred by investigators because of its ease of application and because it does not require any sophisticated calculations. This feature has been our focus while formulating the new scoring system.

The new score (Elsabagh et al.'s scoring system for OLP) included 4 variables: nature of the disease, pain, site, and desquamative gingivitis grading. First, the nature of the disease was divided into the 3 main clinical forms of OLP (reticular, atrophy/erosive, ulcerative), with increasing scores for increasing depth of mucosal destruction. The same major forms were included in other scoring systems suggested by several authors.[6,8,20,21] Some of these systems designated separate scores for each of the 10 or more included mucosal sites, thus complicating calculation.[6,8,20] On the other hand, some authors vary the score according to the size of the lesion, which impedes interobserver reliability,[9,21] whereas others include both drawbacks.[8,14] Such drawbacks were addressed in the new score.

The classification of the nature of the lesion in the new score makes it easier to use, and it was validated by construct validity testing, in which a strong correlation was seen between the lesion nature and the total score and also between the lesion nature and pain (which is one of the most important disease severity predictors, as clarified next).

Including pain as an integral part of the score is of paramount importance. A reduction in pain can be considered the primary goal of OLP management and the main reason for seeking treatment. In addition, treatment that adequately reduces pain is usually considered successful despite lesion persistence.[13] However, many studies have emphasized the importance of pain estimation in disease severity when judging treatment efficacy.[7,8,13,20,22] Several authors used pain and/or subjective scales as separate scoring systems for judging treatment outcomes of OLP,[8,23] and only 1 study combined both clinically reported outcomes and pain estimation into 1 comprehensive score instead of reporting severity as 2 separate scores.[20] It is worth noting that this integration of outcomes, although rare for OLP, is evident in multiple scoring systems of other oral diseases, such as oral mucositis and cicatricial pemphigoid.[20,24-26]

Table IV. Interexaminer agreement (kappa) between the biopsy results (actual disease nature) and Thongprasom's score and the new score (objective disease nature)

	Biopsy				
Thongprasom	*Keratosis*	*Atrophy*	*Erosion*	*Ulceration*	*Total*
Keratosis	1	1	0	0	2 (6.9%)
Atrophy	0	3	5	2	10 (34.5%)
Erosion	0	3	3	11	17 (58.6%)
-	0	0	0	0	0 (0.0%)
	1 (3.4%)	7 (24.1%)	8 (27.6%)	13 (44.8%)	29
Kappa	0.03163				
P	.1920				
95% CI	−0.11 to 0.173				

	Biopsy			
New score	*Keratosis*	*Erosion/atrophy*	*Ulceration*	*Total*
Keratosis	1	1	0	2 (6.9%)
Erosion/atrophy	0	13	2	15 (51.7%)
Ulceration	0	1	11	12 (41.4%)
	1 (3.4%)	15 (51.7%)	13 (44.8%)	29
Kappa	0.75630			
P	<.0001			
95% CI	0.526-0.985			

CI, confidence interval.

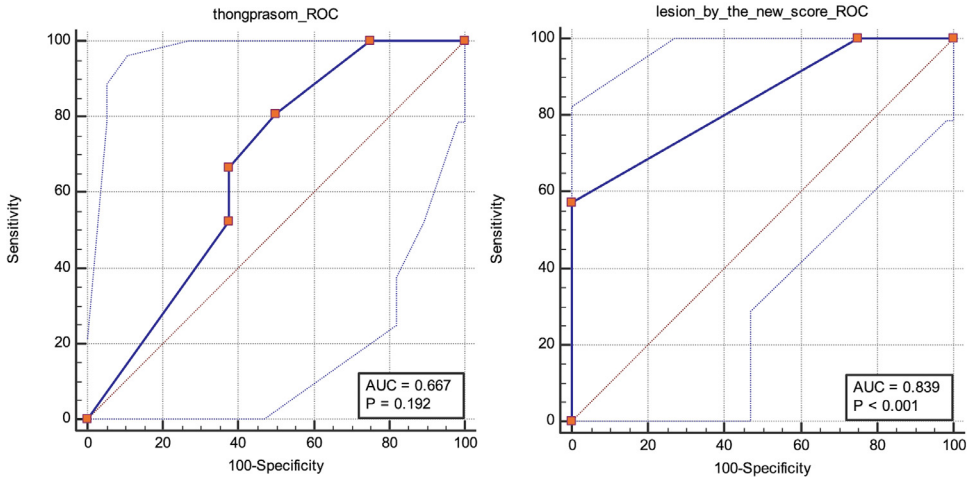

Fig. 3. Receiver operating characteristic curve for both scoring systems.

In fact, because of the diversity of both the objective and subjective parameters involved in various scores, López-Jornet and Camacho-Alonso reported that it is impossible to compare the results obtained from different studies on the treatment of OLP.[15] The compelling need for 1 comprehensive score that combines patient-reported outcomes with clinically reported outcomes has been also suggested by Wang and van der Waal in the context of critical appraisal of more than 22 lichen planus scoring systems.[7] In our study, we attempted to represent the severity of OLP in 1 comprehensive score rather than representing both outcomes as 2 separate scores for the same treatment modality.

In the basic construction of the scoring system, the presence of pain explains the results for concurrent validity, in which a significant strong correlation was found between the new score and NRS (r_s = 0.846), whereas the correlation with Thongprasom's score was weaker (r_s = 0.665). It can be concluded that the new score can provide accurate judgment of pain, therefore excluding the need for the use of a separate pain score (visual analog scale/NRS), as pointed out previously.[8,20,21] This was further confirmed by the strong correlation (0.89) found between pain and the total score that was revealed during construct validity testing.

With regard to site scoring, bilateral buccal mucosal involvement is the most common presentation of the disease. Thus, a score of 0 was suggested for both bilateral and single mucosal site involvement, which represents the least severe disease presentation, and a score of 1 otherwise. Scoring each mucosal site separately is considered by clinicians to be a tedious process that complicates calculations. Furthermore, because some forms of lichen planus (reticular, plaque forms) are asymptomatic yet commonly occur, it was thought that scoring such sites would result in a high total score, providing a false impression of increased disease severity despite the absence of symptoms.

Among the studies currently available in the literature, in an attempt to avoid such a source of complication, some scoring systems completely omitted the consideration of site, which is considered a major drawback in the existing literature.[9,21] On the other hand, such a complexity was demonstrated in a study that compared 2 different scoring systems with detailed site scoring, in which ICC (testing the degree of agreement between raters) decreased with decreasing rater experience.[22] The same issue was highlighted in a study by Escudier et al.[20]

Desquamative gingivitis is a special presentation of OLP and in some cases constitutes the sole manifestation of the disease. In fact, 10% of OLP cases were found to have exclusive gingival involvement,[27] whereas in another study, 36% of desquamative gingivitis cases were found to be caused by OLP.[28] Some authors suggested special scoring systems for OLP-induced desquamative gingivitis that emphasized solicitude for this form.[28,29] Considering these facts, the fourth variable in the new score was selected for the evaluation of the extent of desquamative gingivitis involvement. It was thought that including this item under the site scoring variable would underestimate its importance. Moreover, only the atrophic/erosive form of gingival involvement was considered, because it reflects the severity of disease because of the pain it usually causes.

In this study, the inter- and intraobserver reliability was found to be almost perfect, with an ICC greater than 0.9 among the 3 examiners for both the new scoring system and the one suggested by Thongprasom. This emphasizes the ease of learning and using the new score efficiently with no learning curve, which was an important target to achieve.

In an attempt to prove the validity of the new scoring system, several statistical tests were carried out, including correlations, construct validity testing, and kappa testing for agreement with objective biopsy results.

The new score showed a statistically significant substantial agreement with biopsy results, demonstrating a kappa value of 0.74 ($P < .05$), whereas Thongprasom's score showed nearly no agreement with biopsy results, demonstrating a kappa value of 0.03. This can be attributed to the absence of an ulceration category in Thongprasom's score, in which the highest score was designated to erosive areas of >1 cm^2. In addition, the agreement with biopsy results was definitely compromised by the inclusion of lesion size measured in centimeters. Size inclusion could alter the picture of disease severity, in which atrophic lesions are given a score of 2 and 3 according to size, whereas erosive lesions score 4 and 5. Thus, the failure to distinguish clinically between these 2 lesion types, which is relatively common, would change the final score significantly, especially when using a small-grade scale such as Thongprasom's score. In the newly proposed score, such an error was avoided because atrophy and erosion were merged into 1 entity.

The diagnostic accuracy was much higher for the new scoring system (AUC = 0.83), with a statistical significance that was absent for Thongprasom's score (P = .192, AUC = 0.667). Furthermore, the perfect specificity (100%) of the new score compared with Thongprasom's score (50%) guarantees its capability to discriminate low severity disease (keratosis/atrophy), with very few false-negative cases.

To precisely weigh the effectiveness of the different treatment modalities, the ability of a scoring system to detect minor clinical improvements with treatment is crucial. Clinical sensitivity results demonstrated the ability of the new scoring system to detect clinical improvements in 100% of the cases between baseline and 1-month follow-up, which was significantly different from Thongprasom's scoring system (73.3%). This difference was present but smaller when comparing the 2 follow-up visits and did not reach statistical significance. This could be attributed to the wide-ranging scale of the new score (0-9) in comparison to Thongprasom's score (0-5) and the inclusion of several variables in the new score that enabled it to detect minute improvements during follow-up.

Despite the positive results of this study, which encourage the use of the new scoring system in upcoming clinical trials investigating OLP treatment modalities, this study had several limitations. First, testing the score in further studies that include larger sample sizes is mandatory. Another drawback was the failure to obtain accurate biopsy results for some cases, which were excluded, thus decreasing the sample size available for part of the statistical analysis.

CONCLUSION

The new scoring system could be a useful tool for accurate assesment of the severity of OLP cases in a reproducible manner. It proved to be comprehensive, simple, easy to learn, and easy to apply in a fast, chairside manner without the need for sophisticated calculations. The results of this study showed statistically significant agreement of this score with objective

biopsy results and proved its adequate diagnostic accuracy and clinical sensitivity.

ACKNOWLEDGMENTS

The authors express their gratitude to the staff of the Oral Medicine, Periodontology, Oral Diagnosis and Oral Radiology Department of Alexandria University, Egypt, and to Dr. Hams Hamed for her help in statistical data analysis.

REFERENCES

1. Scully C, Carrozzo M. Oral mucosal disease: lichen planus. *Br J Oral Maxillofac Surg*. 2008;46:15-21.
2. Sugerman PB, Savage NW, Walsh LJ, et al. The pathogenesis of oral lichen planus. *Crit Rev Oral Biol Med*. 2002;13:350-365.
3. Van der Waal I. Oral lichen planus and oral lichenoid lesions; a critical appraisal with emphasis on the diagnostic aspects. *Med Oral Patol Oral Cir Bucal*. 2009;14:E310-E314.
4. Eisen D. The therapy of oral lichen planus. *Crit Rev Oral Biol Med*. 1993;4:141-158.
5. Eisen D, Carrozzo M, Bagan Sebastian J-V, Thongprasom K. Number V oral lichen planus: clinical features and management. *Oral Dis*. 2005;11:338-349.
6. Piboonniyom S-O, Treister N, Pitiphat W, Woo S-B. Scoring system for monitoring oral lichenoid lesions: a preliminary study. *Oral Surg Oral Med Oral Pathol Oral Radiol Endod*. 2005;99:696-703.
7. Wang J, van der Waal I. Disease scoring systems for oral lichen planus; a critical appraisal. *Med Oral Patol Oral Cir Bucal*. 2015;20:e199-e204.
8. Chainani-Wu N, Silverman S Jr, Reingold A, Bostrom A, Lozada-Nur F, Weintraub J. Validation of instruments to measure the symptoms and signs of oral lichen planus. *Oral Surg Oral Med Oral Pathol Oral Radiol Endod*. 2008;105:51-58.
9. Thongprasom K, Luangjarmekorn L, Sererat T, Taweesap W. Relative efficacy of fluocinolone acetonide compared with triamcinolone acetonide in treatment of oral lichen planus. *J Oral Pathol Med*. 1992;21:456-458.
10. Paice JA, Cohen FL. Validity of a verbally administered numeric rating scale to measure cancer pain intensity. *Cancer Nurs*. 1997;20:88-93.
11. Elsabagh HH, Moussa E, Am S, Elsaka RO, Abdelrahman H. Efficacy of melatonin in prevention of radiation induced oral mucositis: a randomized clinical trial. *Oral Dis*. 2020;26(3):566-572.
12. Boonstra AM, Stewart RE, Köke AJA, et al. Cut-off points for mild, moderate, and severe pain on the numeric rating scale for pain in patients with chronic musculoskeletal pain: variability and influence of sex and catastrophizing. *Front Psychol*. 2016;7:1466
13. Lodi G, Carrozzo M, Furness S, Thongprasom K. Interventions for treating oral lichen planus: a systematic review. *Br J Dermatol*. 2012;166:938-947.
14. Zakrzewska JM, Chan ES, Thornhill MH. A systematic review of placebo-controlled randomized clinical trials of treatments used in oral lichen planus. *Br J Dermatol*. 2005;153:336-341.
15. Lopez-Jornet P, Camacho-Alonso F. Clinical assessment of oral lichen planus based on different scales. *Int J Dermatol*. 2010;49:272-275.
16. Qataya PO, Elsayed NM, Elguindy NM, Ahmed Hafiz M, Samy WM. Selenium: a sole treatment for erosive oral lichen planus (randomized controlled clinical trial). *Oral Dis*. 2020;26:789-804.
17. Lavaee F, Shadmanpour M. Comparison of the effect of photodynamic therapy and topical corticosteroid on oral lichen planus lesions. *Oral Dis*. 2019;25:1954-1963.
18. Lajevardi V, Ghodsi SZ, Hallaji Z, Shafiei Z, Aghazadeh N, Akbari Z. Treatment of erosive oral lichen planus with methotrexate. *J Deutsc Dermatol Ges*. 2016;14:286-293.
19. Mutafchieva MZ, Draganova-Filipova MN, Zagorchev PI, Tomov GT. Effects of low level laser therapy on erosive-atrophic oral lichen planus. *Folia Med*. 2018;60:417-424.
20. Escudier M, Ahmed N, Shirlaw P, et al. A scoring system for mucosal disease severity with special reference to oral lichen planus. *Br J Dermatol*. 2007;157:765-770.
21. Kaliakatsou F, Hodgson T, Lewsey J, Hegarty A, Murphy A, Porter S. Management of recalcitrant ulcerative oral lichen planus with topical tacrolimus. *J Am Acad Dermatol*. 2002;46:35-41.
22. Gobbo M, Rupel K, Zoi V, et al. Scoring systems for oral lichen planus used by differently experienced raters. *Med Oral Patol Oral Cir Bucal*. 2017;22:e562-e571.
23. Silverman S, Gorsky M, Lozada-Nur F, Giannotti K. A prospective study of findings and management in 214 patients with oral lichen planus. *Oral Surg Oral Med Oral Pathol*. 1991;72:665-670.
24. World Health Organization. *Handbook for Reporting Results of Cancer Treatment*. WHO; 1979.
25. Ormond M, McParland H, Thakrar P, et al. Validation of an oral disease severity score (ODSS) tool for use in oral mucous membrane pemphigoid. *Br J Dermatol*. 2020;183:78-85.
26. Wee JS, Shirlaw PJ, Challacombe SJ, Setterfield JF. Efficacy of mycophenolate mofetil in severe mucocutaneous lichen planus: a retrospective review of 10 patients. *Br J Dermatol*. 2012;167:36-43.
27. Al-Abeedi F, Aldahish Y, Almotawa Z, Kujan O. The differential diagnosis of desquamative gingivitis: review of the literature and clinical guide for dental undergraduates. *J Int Oral Health*. 2015;7:88-92.
28. Holmstrup P, Schiøtz AW, Westergaard J. Effect of dental plaque control on gingival lichen planus. *Oral Surg Oral Med Oral Pathol*. 1990;69:585-590.
29. Rønbeck BA, Lind PO, Thrane PS. Desquamative gingivitis: preliminary observations with tetracycline treatment. *Oral Surg Oral Med Oral Pathol*. 1990;69:694-697.

Reprint requests:

Hossam H. Elsabagh, BDs, MSc,
Assistant Lecturer of Oral Medicine, Periodontology, Oral Diagnosis and Oral Radiology Department
Faculty of Dentistry, Alexandria University
Champolion St. Azurite
Alexandria 21500
Egypt.
Hossam.hamdy@alexu.edu.eg