Article type : Original Article

Manuscript for British Journal of Dermatology

TITLE:

Validation of an Oral Disease Severity Score (ODSS) tool for use in oral mucous membrane pemphigoid

M Ormond[1*], **H McParland[1*]**, P Thakrar[5], ANA Donaldson[2], M Andiappan[2], RJ Cook[1, 3], ME Escudier[1, 4], J Higham[5], E Hullah[1], R McMillan[6], J Taylor[7], PJ Shirlaw[1], SJ Challacombe[1, 4], JF Setterfield[1, 4, 8].

[1]*Department Oral Medicine, Guy's Hospital, Guy's and St Thomas' NHS Foundation Trust, London, UK*

[2]*Biostatistics and Research Methods Centre, King's College London Faculty of Dentistry, Oral & Craniofacial Sciences, London, UK*

[3]*Centre for Oral Clinical and Translational Sciences, King's College London Faculty of Dentistry, Oral & Craniofacial Sciences, London, UK*

[4]*Centre for Host-Microbiome Interactions, King's College London Faculty of Dentistry, Oral & Craniofacial Sciences, London, UK*

[5]*Department of Oral Medicine, Birmingham Dental Hospital and School of Dentistry, Birmingham, UK*

[6]*Department of Oral Medicine, Eastman Dental Hospital, UCLH / Eastman Dental Institute, UCL, London, UK*

[7]*Department of Oral Medicine, Glasgow Dental Hospital and School, Glasgow, UK*

[8]*St John's Institute of Dermatology, Guy's and St Thomas' NHS Foundation Trust, London, UK*

**these authors contributed equally to this work*

Correspondence to:

Jane Setterfield

E-mail: jane.setterfield@kcl.ac.uk

Department of Oral Medicine, Guy's Hospital, Guy's and St Thomas's NHS Foundation Trust, London, UK

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1111/bjd.18566

No external application for funding was been made.

No conflicts of interest to declare.

What is already known about this topic?

- There are no validated scoring methodologies for oral mucous membrane pemphigoid (MMP). Proposed disease activity scoring tools for MMP include the Mucous Membrane Disease Area Index (MMPDAI) and the Autoimmune Bullous Skin Disorder Intensity Score (ABSIS).
- The Oral Disease Severity Score (ODSS) has been validated for use in oral pemphigus vulgaris (PV). It has been shown to be reliable and sensitive in both lichen planus (LP) and MMP.

What does this study add?

- The ODSS has been shown to be a thorough, sensitive, reproducible, yet quick scoring tool for the assessment of oral involvement in MMP.
- Its versatility for use in oral PV, MMP and LP is an added advantage over other scoring methodologies.

What are the clinical implications of this work?

- We propose that the ODSS be used as a clinical scoring tool for **monitoring** disease activity in **oral** MMP **in clinical practice** as well as for **use in** multicentre studies.

ABSTRACT

Background

Mucous membrane pemphigoid (MMP) is a rare autoimmune bullous disease predominantly affecting the oral mucosa. Optimal management relies upon thorough clinical assessment and documentation at each visit.

Objectives

The primary aim of this study was to validate the Oral Disease Severity Score (ODSS) for the assessment of oral involvement in MMP. We also compared its inter- and intra-observer reliability with the oral parts of the Mucous Membrane Pemphigoid Disease Area Index

(MMPDAI), Autoimmune Bullous Skin Disorder Intensity Score (ABSIS) and Physician's Global Assessment (PGA).

Methods

Fifteen patients with mild to moderately severe oral MMP were scored for disease severity by 10 oral medicine clinicians from four UK centres using the ODSS, the oral sections of MMPDAI and ABSIS and PGA. Two clinicians re-scored all patients after two-hours.

Results

Inter-observer reliability for ODSS total score intra-class correlation coefficient (ICC) was 0.97; MMPDAI activity 0.59 and damage 0.15; ABSIS total 0.84; and PGA 0.72. Intra-observer ICC for ODSS total was 0.97 and 0.93; MMPDAI activity 0.93 and 0.70 and damage 0.93 and 0.79; ABSIS total 0.99 and 0.94; and PGA 0.92 and 0.94. Convergent validity between ODSS and MMPDAI was good (correlation coefficient 0.88). The mean±SD time (seconds) for completion, ODSS 93±31 s, MMPDAI 102±24 s and ABSIS involvement 71±18 s. The PGA was <5 s.

Conclusion

This study has validated the ODSS for the assessment of oral MMP. It has shown superior inter-observer reliability than MMPDAI, ABSIS and PGA and intra-observer reliability than MMPDAI. It is quick and easy to perform.

INTRODUCTION

Mucous membrane pemphigoid (MMP) is a heterogeneous disease with a wide spectrum of clinical and immunopathological presentations. Affected sites include the oral mucosa, eye, skin, genitalia, larynx, oesophagus and nasopharynx.[1] Oral manifestations of MMP include painful desquamative gingivitis, blistering and ulceration of the palate, buccal mucosae and oropharynx with occasional lesions seen on the floor of mouth, tongue and lips. There appear to be distinct oral phenotypes.[1] Approximately a third of patients will experience skin lesions which are typically seen on the head and neck area though occasionally also affecting the limbs.[2] In a small subgroup of patients it may present with a more widespread cutaneous blistering condition akin to bullous pemphigoid (BP), but the oral lesions are usually more

prominent from the outset than would be expected in BP and are a clue to the diagnosis of MMP. The disease can have a long course and rarely remits without treatment. A wide variety of therapies have been used to treat MMP but there are no large placebo-controlled randomised controlled studies, and there are few studies utilising a standardised scoring methodology to quantify objective improvement with systemic treatment.[3]

As the mouth and eyes are the two most frequently involved sites in MMP, it is paramount that these are assessed with the most accurate, reproducible and validated methods available. In addition, to be valuable, a scoring tool should also be feasible, sensitive to change, and have external validity.[4] To date there have been no validated scoring methodologies for any site in MMP.[5] For ocular disease the most frequently used prospective scoring system is the Foster-Tauber tool in which the presence of subconjunctival scarring, together with qualitative assessments of the extent of both forniceal fore-shortening and symblepharon, are combined to create a four stage alpha-numeric measure of the extent of conjunctival scarring.[6] However further ocular MMP scoring methodologies including assessments of the severity of inflammation using image based grading and quantitative measurements of forniceal shortening are under evaluation.[7,8,9] In 1998 we published a multi-site scoring methodology for MMP. It was used to assess disease severity alongside establishing biomarkers for more severely affected patients.[1] Each affected site was formally assessed by the relevant specialist e.g. in dermatology, oral medicine, otolaryngology or ophthalmology and the methodology was devised with their input. This scoring methodology was later used in further MMP cohort studies.[10] However, for more accurate sequential assessment, a more detailed oral severity score was needed.

The Oral Disease Severity Score (ODSS) was devised by the oral medicine group at **Guy's Hospital** as part of a strategy to develop a comprehensive scoring system to record outcomes for oral mucosal diseases.[11] It was developed from a multi-site methodology previously published for MMP.[1] ODSS records the presence of lesions and degree of activity at multiple oral sites and includes a subjective assessment of the patient's oral pain over the preceding week. Previous studies have demonstrated that the ODSS is a reliable and sensitive tool in both oral lichen planus (LP) and MMP.[11,12] It has been recently validated for use in oral PV[13] and has been shown to be useful to assess therapeutic response over time in both severe mucosal LP and PV.[14,15]

In 2012, an International panel of experts, predominantly dermatologists, proposed a new scoring system, the Mucous Membrane Pemphigoid Disease Area Index (MMPDAI).[16] This was adapted from the validated Pemphigus Disease Area Index (PDAI) and the Bullous Pemphigoid Disease Area Index (BPDAI)[17,18,19], and is a multisite scoring methodology that is yet to be validated. A further tool advocated for potential use in MMP is the Autoimmune Bullous Skin Disorder Intensity Score (ABSIS), which has been validated for pemphigus vulgaris (PV) but not for MMP .[20]

The primary aim of our study was to validate the ODSS for use in MMP by investigating inter- and intra-observer reliability and ease of use. A secondary aim was to compare its inter- and intra-observer reliability and ease of use with the oral parts of the MMPDAI, ABSIS and the Physician's Global Assessment (PGA).

PATIENTS & METHODS

Research ethics approval was obtained (REC15/ES/0038). The study was conducted over one day within the Department of Oral Medicine at Guy's Hospital, London.

Patients

Fifteen patients (aged 31-79) with a confirmed diagnosis of predominantly oral MMP (based on clinical findings, histopathology and direct immunofluorescence) were recruited consecutively from the outpatient clinic of the Departments of Oral Medicine and Dermatology at Guy's Hospital, London. The visit replaced one of their routine follow-up appointments. All patients had mild to moderately severe oral lesions. Fourteen patients were taking systemic treatment: eight sulphapyridine, two dapsone, one azathioprine, one mycophenolate mofetil, one prednisolone plus sulphapyridine and one prednisolone plus mycophenolate mofetil. One patient was using topical treatment only (fluticasone propionate nasules).

Physicians

Ten clinicians experienced in diagnosing and managing patients with MMP were included from four oral medicine centres in the UK. All were oral medicine specialists; six were medically and dentally qualified, one of whom was a practicing dermatologist. Patients were scored using the ODSS, the oral parts of the MMPDAI and ABSIS and the PGA. All physicians were familiar with the PGA, five were experienced in using ODSS and five were

not. None of the clinicians routinely used either the MMPDAI or ABSIS. Prior to the study, a set of training slides demonstrating the ODSS system, MMPDAI, ABSIS and PGA was sent to all clinicians. On the study day the chief investigator met with all the clinicians for a detailed discussion of methodologies using clinical slides as examples. All clinicians examined and scored each patient once, and two clinicians examined all patients twice with a two-hour interval to reduce recall. All physicians were asked for feedback regarding the scoring tools in relation to ease of use. An assistant recorded the scores in random order and the time taken for each methodology. Twelve sets of scores were recorded for each patient.

Oral Disease Severity Score

The ODSS is a comprehensive oral scoring system detailed in Figure 1 for MMP and previously validated for oral PV.[13] It has been used as an outcome measure in the assessment of therapeutic efficacy in oral LP and PV.[11,14,15] In the ODSS the oral cavity is divided into 17 sites weighted 0-2 according to the area of possible involvement. These sites are the outer / inner lips, buccal mucosae right / left, soft palate right / left, hard palate right / left, dorsum of tongue right / left, ventrolateral tongue right / left, floor of mouth right / left, oropharynx right / left and the gingivae (divided into 6 segments). As detailed in Figure 1, a score of 2 corresponds to >50% of the buccal mucosa on one side being affected, or bilateral involvement of the dorsum of tongue, floor of mouth, hard and soft palate or oropharynx. Each unit of site is then allocated an activity score which ranges from 0-3 (no activity = 0, mild inflammation – erythema or healing areas = 1, prominent erythema = 2, and blistering or ulceration = 3) (Fig 2a, 2b and 2c). For a specific area e.g. buccal mucosa, the activity for each unit of site is added together. The third component is a pain score, which is subjective on a scale of 0-10 provided by the patient as an average for the preceding week. The three components are summed to give a total score with a theoretical maximum of 106; however greater than 95% of patients would be expected to have scores in the range from 0-60 representing a clinical range from remission to severe disease.

Mucous Membrane Pemphigoid Disease Area Index

The MMPDAI was designed through consensus by a group of international experts in autoimmune bullous disease and is currently awaiting validation.[16] Only the section relating to the oral cavity was used in our study. The oral cavity is divided into seven areas, the buccal mucosa, palate, upper and lower gingiva, tongue / floor of mouth, labial mucosa, and posterior pharynx. Scores are recorded in two columns, to separate active ulcers and blisters

from post inflammatory changes and scarring (damage). Scores are allocated based on number and size of lesions with a total possible score of 70 points for mucosal activity and 7 points for disease damage. The **activity and damage scores are not added together**.

Autoimmune Bullous Skin Disorder Intensity Score

The ABSIS scoring tool was developed by Pfutze[20], initially for use in PV to score both skin and mucous membrane involvement. In our study, only the oral part of the tool was used. Oral involvement is scored by recording the presence or absence of lesions in 11 sites in the oral cavity, with each site scoring 0 or 1 and a maximum of 11. The oral cavity is divided into upper gingivae, lower gingivae, upper lip, lower lip, right buccal mucosa, left buccal mucosa, tongue, floor of mouth, hard palate, soft palate and pharynx. A subjective severity scale based on the patient's assessment of discomfort during eating and drinking a range of foods is included with maximum score of 45. The two components are summed to provide a total score for oral severity.

Physician Global Assessment

The PGA is a 10-point analogue scale in which clinicians make a judgement on the overall health of, in this case, the oral mucosa. The scorer rates the mucosa from 0 = perfect health to 10 = worst mucosal disease imaginable. PGA has previously been used as an outcome measure in dermatological conditions[21,22,23], and has been validated for oral PV.[13] It has not been validated for use in other autoimmune bullous disorders.

Time for scoring methodology completion

An independent assistant used a stopwatch to record the time taken (in seconds) by each clinician to obtain a disease severity score for each scoring tool except the PGA which took <5 seconds and was therefore not timed. **ODSS time included the subjective pain component whereas the ABSIS timing did not include the subjective severity score.**

Statistical Methods

Inter-observer reliability was assessed by 10 clinicians (observers) scoring all patients with each of the four scoring tools. A sample size of 15 subjects was required to achieve intra-class correlations (ICC) of 0.77 for the inter-observer reliability.

Intra-observer reliability was tested with two replications per subject (as per test-retest) with a minimum of two-hours between scores to minimise the risk of recall. Since the involvement was more onerous (burden, time or money resources, etc.) for the rater than for the subject, taking rater as fixed in the factorial design was more efficient. We fixed the number of raters to two and found the sample size required in terms of the number of subjects (which are assumed to be a random sample from the population of subjects). With both raters performing two replications in each methodology in all the subjects, a total of 15 subjects provided 80% power to detect an ICC difference of 0.50 (relative to a null value of 0.20). Anticipating an ICC of 0.85, 15 subjects with two replications will yield a width of 0.30 in the 95% confidence interval (CI).

Multilevel models were used to quantify inter- and intra-observer reliabilities of the continuous measures. Assessment for the level of agreement in terms of the ICC for ordinal or continuous measures followed well-established benchmark limits (Fleiss and Altman's benchmark scales).[24,25] Landis-Koch's benchmark values were followed when Kappa coefficients were used for categorical outcomes.[26] In all cases, for more rigour, in addition to the point estimate, we took into account the lower bound of the 95% confidence interval.[24] Convergent validity was calculated using the Spearman rank correlation coefficient.

RESULTS

15 patients (f5:m10) with confirmed MMP were included. The mean±SD age was 65±11.8 years (range 31-79).

The distribution of scores:

The mean±SD total ODSS score was 25.6±10.9, (range 6-56); median 25 with interquartile range [IQR] 17-32.8; reflecting mild-moderately severe disease. The mean MMPDAI activity score was 7.2±6.1, (range 0-31); median 6 [2-11] and mean damage was 1.6±1.4, (range 0-6); median 1.5 [0-3]. The mean ABSIS total score was 12.1±10.5 (range 0-36) and median 10 [2-20]. The mean score for involvement was 3.2 ±1.5, median 3 [2-4]. For disease severity the mean was 9.9±9.6, median 8 [0-17.5]. The mean PGA score was 4.1±2.2, (range 1-9), median 4 [2-6] (Table 1).

Reliability

Inter-observer reliability

The ICC (95% CI) for the ODSS total was 0.97 (0.94-0.99), ODSS site 0.73 (0.57-0.88), ODSS activity 0.82 (0.71-0.94) and ODSS pain 0.81 (0.68-0.93). For MMPDAI the ICC for activity was 0.59 (0.39-0.79) and for damage 0.15 (0.08-0.30). For ABSIS the ICC total score was 0.84 (0.74-0.95), for involvement 0.74 (0.59-0.89) and for patient reported severity 0.87 (0.78-0.96). Finally, the ICC for the PGA was 0.72 (0.56-0.88) (Table 1). Table 1 also includes benchmark values following Fleiss and Altman's benchmark scales.[24,25]

Test-retest reliability

Intra-observer agreement between initial scoring and rescoring of the same patients demonstrated an ICC for ODSS total of 0.97 (0.93-1.00) and 0.93 (0.86-0.99), site 0.95 (0.90-1.00) and 0.93 (0.85-1.00), activity 0.95 (0.90-1.00) and 0.94 (0.88-1.00) and pain 0.97 (0.94-1.00) and 0.59 (0.25-0.93). The ICCs for MMPDAI activity were 0.93 (0.87-1.00) and 0.70 (0.44-0.96) and damage 0.93 (0.85-1.00) and 0.79 (0.60-0.98). For ABSIS total, ICCs were 0.99 (0.98-1.00) and 0.94 (0.89-1.00), involvement 0.98 (0.97-1.00) and 0.90 (0.79-1.00) and severity 0.99 (0.97-1.00) and 0.94 (0.99-1.00). The PGA ICCs were 0.92 (0.84-1.00) and 0.94 (0.89-1.00) (Table 2).

Convergent Validity

There was good correlation between the MMPDAI activity and the ODSS total score (0.884, $p<0.001$), ODSS total with ABSIS (0.797, $p<0.001$) and MMPDAI activity and ABSIS (0.791, $p<0.001$) (Table 3).

Time for Completion

The mean time to obtain a disease severity score using ODSS (total) was 93±31 s, MMPDAI (activity and damage) 102±24 s and ABSIS (involvement) 71±18 s.

DISCUSSION

This study has shown that the ODSS is a valid scoring method for assessment of oral disease severity in MMP with higher inter- and intra-observer reliability than MMPDAI and higher inter-observer reliability than ABSIS and the PGA. **The study has also validated the oral components of the MMPDAI and ABSIS for use in MMP.**

The methodologies were compared using standardised benchmark scales. The inter-observer score for ODSS total was classified as excellent in contrast to the MMPDAI activity score which was fair / moderate and the MMPDAI damage score which was poor. For ABSIS the total score was good / substantial and for PGA moderate / good. Thus, this study has shown that the ODSS was the more reliable as a scoring tool among these clinicians than MMPDAI, ABSIS or the PGA.

The sample size was calculated in advance using appropriate statistical calculations in order to achieve an intra-class correlation of 0.77, which is considered more than adequate to allow assessment of reliability. As MMP is a rare disease combined with the requirement for all participants to attend on one day, increasing the sample size would have been practically difficult.

We sought statistical advice on how to undertake the intra-observer methodology. By using two clinicians to score each patient twice, 30 scores were produced compared with 20 if each clinician had rescored one patient. We retested with a minimum two-hour interval to reduce recall bias. While a longer interval (24-48 hours) might have further reduced recall, this might also have been associated with subtle changes in disease activity and of practical relevance patients would have required two visits adding an extra burden. **Although 5 out of the 10 investigators had experience of the ODSS, with the potential for possible bias in the data, there was no difference detected in the reliability of the tool between these groups.** For intra-observer reliability the scores for ODSS total, ABSIS total and PGA were classified as excellent. For MMPDAI activity and damage, benchmark ratings were good / substantial. Thus, the ODSS had similar or better intra-observer scores than the other scoring tools examined.

Accuracy of any scoring system is best assessed against a gold standard. However, for MMP no gold standard exists, as there are no validated scoring methods with which to compare. There was good convergent validity between the oral part of MMPDAI activity and ODSS total, ODSS total and ABSIS as well as MMPDAI and ABSIS. The validity of a tool also depends on how well the variables in the study represent the phenomenon of interest (construct validity). **Whilst construct validity was not the aim of this study nor addressed formally, the variables in the ODSS contain objective measures of disease activity and severity as well as including patient subjective data allowing for a comprehensive appraisal of mucosal disease.** [11] **Construct validity may be an area for further**

investigation in the future. These outcome measures were identified as important by the World Workshop in Oral Medicine VI.[27] The scores can be reported separately giving an accurate assessment by both clinician and a patient reported outcome. **It should be noted that although ODSS contains a subjective element it does not replace a quality of life measure. The granularity of the ODSS scoring system allows this component to be assessed separately from the other aspects of disease severity. In order to capture patient experience ODSS, MMPDAI and ABSIS can be combined with patient reported outcome measures**. The ODSS has been externally evaluated for use in MMP[10] and validated for use in PV.[13]

In terms of the practical use of these methodologies in the clinical setting, the average time taken to complete each of the scoring tools was less than two minutes (ODSS 93 s, MMPDAI 116 s and ABSIS 71 s), which we consider feasible for use on routine clinics. The PGA typically takes <5 seconds and was therefore not timed. Since there was no significant difference in the time taken for those familiar with ODSS and other clinicians, we extrapolate that these mean differences reflect ease of use. However, all were **felt to be** practical **for use** in the clinical setting.

Clinicians were asked to comment on their experience of using each scoring tool. They reported that the ODSS was easy to use and most accurately recorded the extent of oral disease in MMP, recording more detailed activity across more sites, thereby having a wider scale. ABSIS was the easiest tool to assess activity recording blisters or ulcers at only 11 sites. However, there was loss in sensitivity as erythema was not included although present in pre-blistering lesions or stable inflammation. The substantial subjective component of ABSIS requires the patient to report symptoms with food types and the results showed differing answers depending upon how the questions were put to the patient. Clinicians felt the ABSIS scoring system was weighted too strongly on this subjective component. The PGA, while simple and very quick, was felt to offer little information regarding the objective oral involvement of MMP so its potential for sequential monitoring of disease was limited.

The MMPDAI is the only scoring tool to include an activity and damage component. For the mouth, it requires lesions to be an ulcer or a blister to be active while scarring, erythema and post-inflammatory changes are descriptors for damage. Inclusion of a damage score for multisite MMP is valuable albeit that scarring is an unusual finding in

the oral mucosa. However, the omission of erythema as a descriptor for activity precludes those lesions that are stable but erythematous and therefore in the oral mucosa, inflamed. **The authors of the MMPDAI have subsequently acknowledged that the term 'active erythema' was missing as a descriptor for the proof of the 2015 paper and an erratum has been submitted.** In contrast to PV, gingival erythema in MMP may last months or years, is usually symptomatic, requires active monitoring and often systemic therapy. Thus, oral physicians unanimously felt this should more correctly represent activity rather than damage and may in part have accounted for the longer time taken to complete the scoring. Clinicians also reported difficulty in assessing the level of activity where there was localised or patchy gingival ulceration, finding the 10-point scale more difficult to adapt to the gingiva.

While the MMPDAI is intended for use by dermatologists in 'mild MMP' but not in patients with severe ocular or laryngeal disease, patients with oral lesions occur in both groups. Furthermore, those with mild disease will present to a range of specialists and therefore a scoring methodology that is applicable to all patients and all clinicians is arguably more useful. As neither oral physicians nor dermatologists are skilled in ocular assessment, and disease progression does not always manifest as erythema, a total score that includes oral and ocular specific methods but with additional sites assessed would be optimal.

An ideal scoring system should be able to assess a patient's baseline severity to aid assessment, monitor progress longitudinally throughout treatment and detect relapse.[4] The ODSS has been used in this way in our department for over 10 years for oral MMP, PV and LP. It has been shown to be valuable tool in longitudinal studies in oral LP and PV[14,15] and has revealed distinct clinical phenotypes in MMP, which appear to be stable over long periods of time.

This study has demonstrated the value of the ODSS for assessment of disease severity in MMP and by assessing **activity at 17** oral sites **it provides** the potential to **accurately** monitor response to treatment. It is easy to use and quick to learn and is designed for use in oral medicine, dermatology and other relevant clinical specialities. Its additional versatility for use in PV and LP is an added advantage over other scoring methodologies. We propose that this scoring tool would be useful for recording sequential disease activity **routinely** in the clinic as well as in future multicentre studies.

REFERENCES

1. Setterfield J, Shirlaw PJ, Kerr-Muir M *et al.* Mucous membrane pemphigoid: a dual circulating antibody response with IgG and IgA signifies a more severe and persistent disease. *Br J Dermatol* 1998; 138:602-610.

2. Ahmed AR, Kurgis BS, Rogers RS. Cicatricial pemphigoid. *J Am Acad Dermatol* 1991; 24:987-1001.

3. Taylor J, McMillan R, Shephard M *et al*. World Workshop on Oral Medicine VI: a systematic review of the treatment of mucous membrane pemphigoid. *Oral Surg Oral Med Oral Pathol Oral Radiol* 2015; 120:161-171.

4. Bastuji-Garin S, Sbidian E. How to validate outcome instruments for pemphigus. *J Invest Dermatol* 2009; 129:2328-2330.

5. Lee BWH, Tan JCK, Radjenovic M, Coroneo MT, Murrell DF. A review of scoring systems for ocular involvement in chronic cutaneous bullous disease. *Orphanet J Rare Dis* 2018; 13:83.

6. Tauber J, Jabbur N, Foster CS. Improved detection of disease progression in ocular cicatricial pemphigoid. *Cornea* 1992; 11:446–51.

7. Ong HS, Setterfield JF, Minassian DC, Dart JK, Mucous Membrane Pemphigoid Study Group 2009-2014. Mucous Membrane Pemphigoid with Ocular Involvement: The Clinical Phenotype and Its Relationship to Direct Immunofluorescence Findings. *Ophthalmology* 2018; 125(4):496-504.

8. Williams GP, Saw VP, Saeed T, et al. Validation of a fornix depth measurer: a putative tool for the assessment of progressive cicatrising conjunctivitis. *Br J Ophthalmol* 2011; 95(6): 842-847.

9. Dart JK. The 2016 Bowman Lecture Conjunctival curses: scarring conjunctivitis 30 years on. *Eye (Lond)* 2017; 31:301-332.

10. Cozzani E, Di Zenzo G, Calabresi I *et al*. Autoantibody Profile of a Cohort of 78 Italian Patients with Mucous Membrane Pemphigoid: Correlation Between Reactivity Profile and Clinical Involvement. *Acta Derm Venereol* 2016; 96:768–773.

11. 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.

12. Reeves GMB, Lloyd M, Rajlawat BP *et al*. Ocular and oral grading of mucous membrane pemphigoid. *Graefes Arch Clin Exp Ophthalmol* 2012; 250:611-618.

13. Ormond M, McParland H, Donaldson ANA *et al*. An Oral Disease Severity Score validated for use in oral pemphigus vulgaris. *Br J Dermatol* 2018; 179:872-81.

14. Wee J, 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.

15. Greenblatt DT, Benton EC, Groves RW, Setterfield JF. Crescendo response to Rituximab in oral pemphigus vulgaris: a case with 7-year follow-up. *Clin Exp Dermatol* 2016; 41:529-32.

16. Murrell DF, Marinovic B, Caux F *et al*. Definitions and outcome measures for mucous membrane pemphigoid: Recommendations of an international panel of experts. *J Am Acad Dermatol* 2015; 72:168-74.

17. Murrell DF, Dick S, Ahmed AR *et al*. Consensus statement on definitions of disease, end points and therapeutic response for pemphigus. *J Am Acad Dermatol* 2008; 58:1043-6.

18. Rosenbach M, Murrell D, Bystryn JC *et al*. Reliability and Convergent Validity of Two Outcome Instruments for Pemphigus. *J Invest Dermatol*. 2009; 129:2404-2410.

19. Wijayanti A, Zhao CY, Boettiger D *et al*. The Reliability, Validity and Responsiveness of Two Disease Scores (BPDAI and ABSIS) for Bullous Pemphigoid: Which One to Use? *Acta Derm Venereol* 2017; 97:24-31.

20. Pfütze M, Niedermeier A, Hertl M, Eming R. Introducing a novel Autoimmune Bullous Skin Disorder Intensity Score (ABSIS) in pemphigus. *Eur J Dermatol* 2007; 17:4-11.

21. Langley RG, Ellis CN. Evaluating psoriasis with psoriasis area and severity index, psoriasis global assessment, and lattice system physician's global assessment. *J Am Acad Dermatol* 2004; 51:563–9.

22. Heald P, Mehlmauer M, Martin AG *et al*. Topical bexarotene therapy for patients with refractory or persistent early-stage cutaneous T-cell lymphoma: results of the phase III clinical trial. *J Am Acad Dermatol* 2003; 49:801-15.

23. Guzzo CA, Weiss JS, Mogavero HS *et al*. A review of two controlled multicenter trials comparing 0.05% halobetasol propionate ointment to its vehicle in the treatment of chronic eczematous dermatoses. *J Am Acad Dermatol* 1991; 25:1179–83.

24. Shrout PE, Fleiss JL. Intraclass correlations uses in assessing rater reliability. *Psychol Bull* 1979; 86(2):420-8.

25. Altman DG. *Practical statistics for medical research*. Chapman and Hall: London, 1991.

26. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977; 33:159-74.

27. Ni Riordain R, Shirlaw P, Alajbeg I *et al*. World Workshop on Oral Medicine VI: Patient-reported outcome measures and oral mucosal disease: current status and future direction. *Oral Surg Oral Med Oral Pathol Oral Radiol* 2015; 120:161-71. e20.

Table 1. Scores and inter-observer reliability for each of the disease severity scoring systems, and their individual components

ODSS, Oral Disease Severity Score; MMPDAI, Mucous Membrane Pemphigoid Disease Activity Index; ABSIS, Autoimmune Bullous Skin Disorder Intensity Score; PGA, Physician's Global Assessment; SD, standard deviation; IQR, interquartile range; ICC, intra-class correlation coefficient; CI, confidence interval

*Assessment for the level of agreement in terms of the intra-class correlation coefficients followed Fleiss and Altman's benchmark scales[22, 23]

Scoring system	Range	Mean ±SD Median (IQR)	Inter-observer ICC (95% CI)	Overall benchmark values*
ODSS Site	0-10	3.2±2.4 3 (2-4.5)	0.73 (0.57-0.88)	Moderate / Good
ODSS Activity	3-37	15.0±7.4 15 (9-19)	0.82 (0.71-0.94)	Good / Substantial
ODSS Pain	0-10	3.2±2.4 3 (2-4.5)	0.81 (0.68-0.93)	Good / Substantial
ODSS Total (0-106)	6-56	25.6±10.9 25 (17-32.75)	0.97 (0.94-0.99)	Excellent
MMPDAI Activity (0-70)	0-31	7.2±6.1 6 (2-11)	0.59 (0.39-0.79)	Fair / Moderate
MMPDAI Damage (0-12)	0-6	1.6±1.4 1.5 (0-3)	0.15 (0.08-0.30)	Poor
ABSIS Involvement	0-8	3.2±1.5 3 (2-4)	0.74 (0.59-0.89)	Moderate / Good
ABSIS Severity	0-32	9.9±9.6 8 (0-17.5)	0.87 (0.78-0.96)	Good / Substantial
ABSIS Total (0-56)	0-36	12.1±10.5 10 (2-20)	0.84 (0.74-0.95)	Good / Substantial
PGA (0-10)	1-9	4.1±2.2 4 (2-6)	0.72 (0.56-0.88)	Moderate / Good

Table 2. Within observer (intra-observer) reliability data for each scoring methodology

ODSS, Oral Disease Severity Score; MMPDAI, Mucous Membrane Pemphigoid Disease Area Index; ABSIS, Autoimmune Bullous Skin Disorder Intensity Score; PGA, Physician's Global Assessment; SD, standard deviation; IQR, interquartile range; ICC, intra-class correlation coefficient; CI, confidence interval

*Assessment for the level of agreement in terms of the intra-class correlation coefficients followed Fleiss and Altman's benchmark scales[22, 23]

Scoring system	Observer 1 ICC (95% CI)	Observer 2 ICC (95% CI)	P-value	Overall benchmark values*
ODSS Site	0.95 (0.90-1.00)	0.93 (0.85-1.00)	0.26	Excellent
ODSS Activity	0.95 (0.90-1.00)	0.94 (0.88-1.00)	0.39	Excellent
ODSS Pain	0.97 (0.94-1.00)	0.59 (0.25-0.93)	0.35	Moderate / Good
ODSS Total	0.97 (0.93-1.00)	0.93 (0.86-0.99)	0.75	Excellent
MMPDAI Activity	0.93 (0.87-1.00)	0.70 (0.44-0.96)	0.30	Good / Substantial
MMPDAI Damage	0.93 (0.85-1.00)	0.79 (0.60-0.98)	0.28	Good / Substantial
ABSIS Involvement	0.98 (0.97-1.00)	0.90 (0.79-1.00)	0.80	Excellent
ABSIS Severity	0.99 (0.97-1.00)	0.94 (0.99-1.00)	0.57	Excellent
ABSIS Total	0.99 (0.98-1.00)	0.94 (0.89-1.00)	0.55	Excellent
PGA	0.92 (0.84-1.00)	0.94 (0.89-1.00)	0.05	Excellent

Table 3. Convergent validity of disease severity scoring systems

ODSS, Oral Disease Severity Score; MMPDAI, Mucous Membrane Pemphigoid Disease Area Index; ABSIS, Autoimmune Bullous Skin Disorder Intensity Score

Scoring system	Correlation coeffficient	P-value
ODSS total & MMPDAI activity	0.884	<0.0001
ODSS total & ABSIS total	0.797	<0.001
MMPDAI activity & ABSIS total	0.791	<0.0001

Figure 1.

Oral Disease Severity Score (ODSS)

Figure 2a.

Upper central gingivae indicating mild erythema (site score = 1, activity 1)

Figure 2b.

Upper central gingivae indicating marked erythema (site score = 1, activity = 2)

Figure 2c.

Hard palate demonstrating areas of ulceration bilaterally (site score = 2, activity = 3+3 = 6)

Site	Site Score	Activity Score / Unit of Site (0-3)*
Outer lips (1)		
Inner lips (1)		
R Buccal mucosa (1 or 2)		
L Buccal mucosa (1 or 2)		
Gingivae (1 each segment)		
Lower R (from 1st premolar)		
Lower central (canine to canine)		
Lower L (from 1st premolar)		
Upper R (from 1st premolar)		
Upper central (canine to canine)		
Upper L (from 1st premolar)		
Dorsum of tongue (1 or 2)		
R Ventral tongue (1)		
L Ventral tongue (1)		
Floor of mouth (1 or 2)		
Hard palate (1 or 2)		
Soft palate (1 or 2)		
Oropharynx (1 or 2)		
Total		

Total Score = Site Score + Activity Score + Pain Score (1-10) (Maximum 106)

Site Score

0 if no lesion
1 if lesion

For the buccal mucosa:

1 if less than 50% of area affected
2 if greater than 50% of area affected

For the dorsum of tongue, floor of mouth, hard or soft palate or oropharynx:

1 unilateral
2 bilateral

Activity Score

1 mild erythema
2 marked erythema without erosion
3 erosion or ulceration

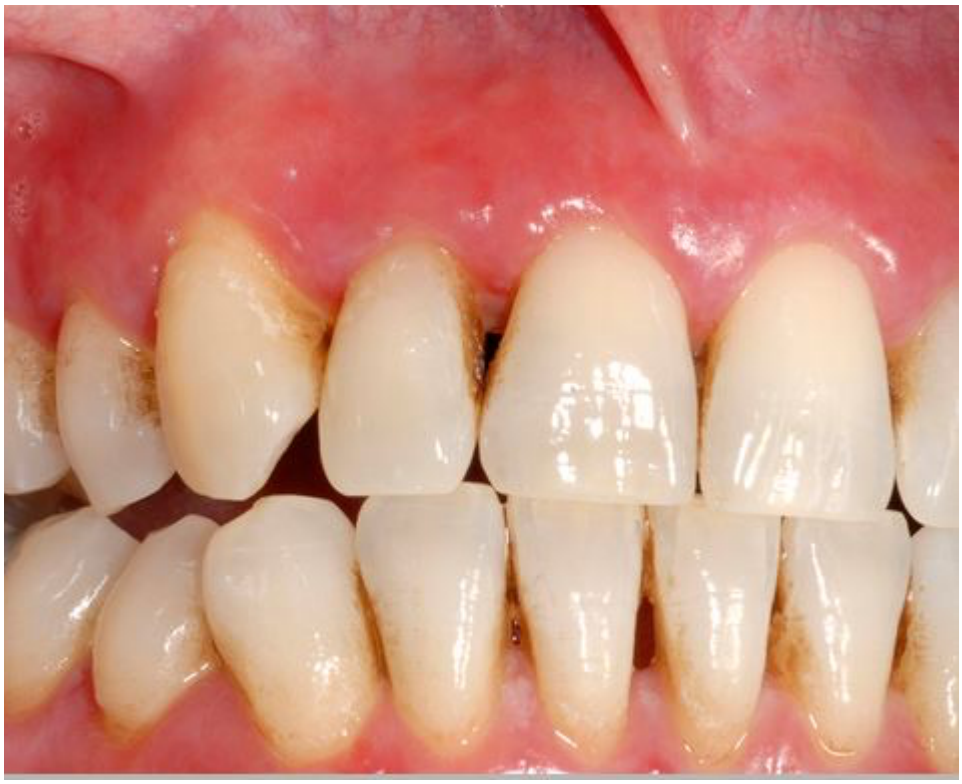
*where a site has a score of 2, each site unit is allocated an activity score, which are then added together

Pain Score

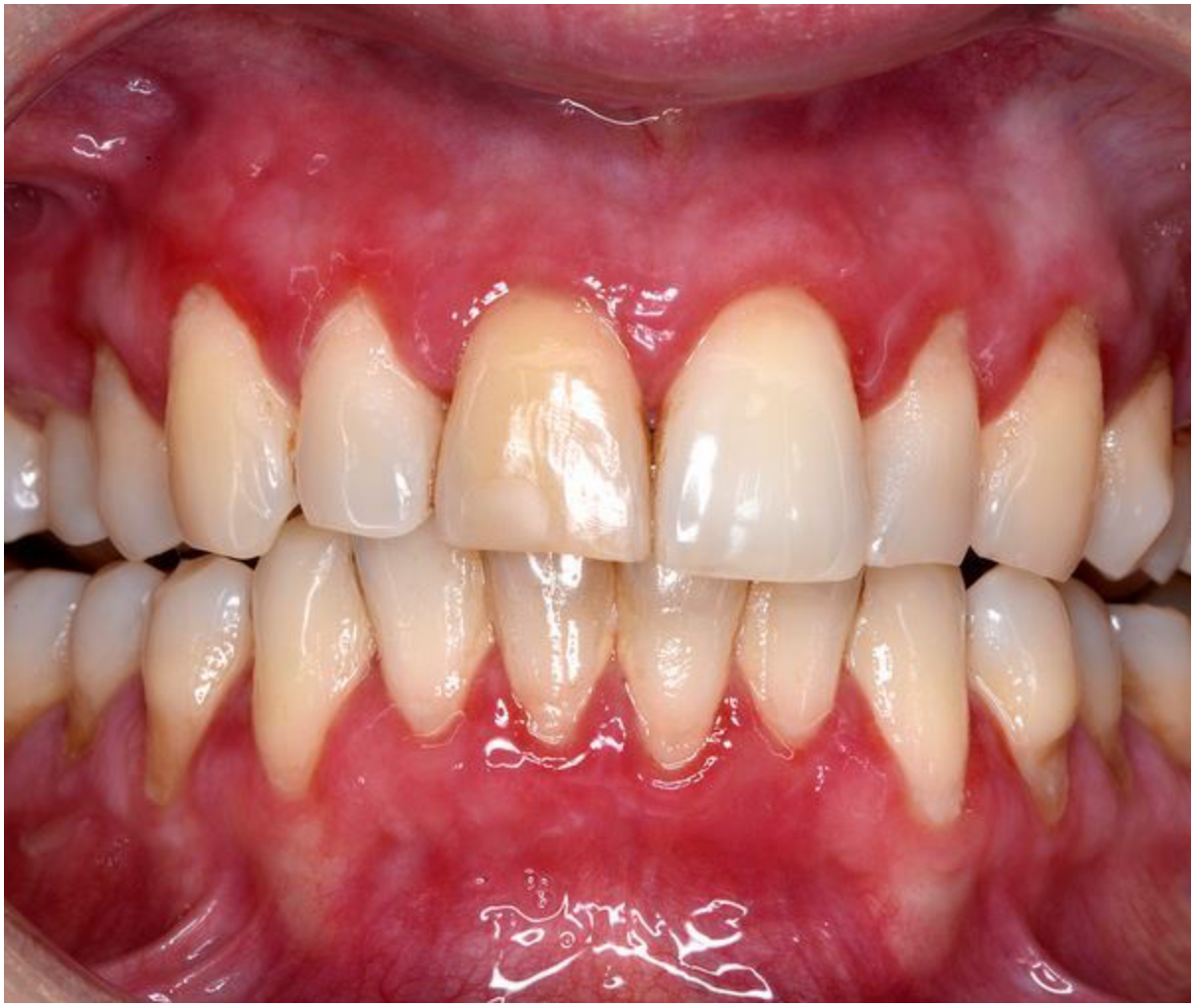
Analogue scale from 0 (no discomfort) to 10 (the most severe pain they have encountered with this condition so far)

The patient is asked to provide a score reflecting their pain / discomfort as an average of the preceding week

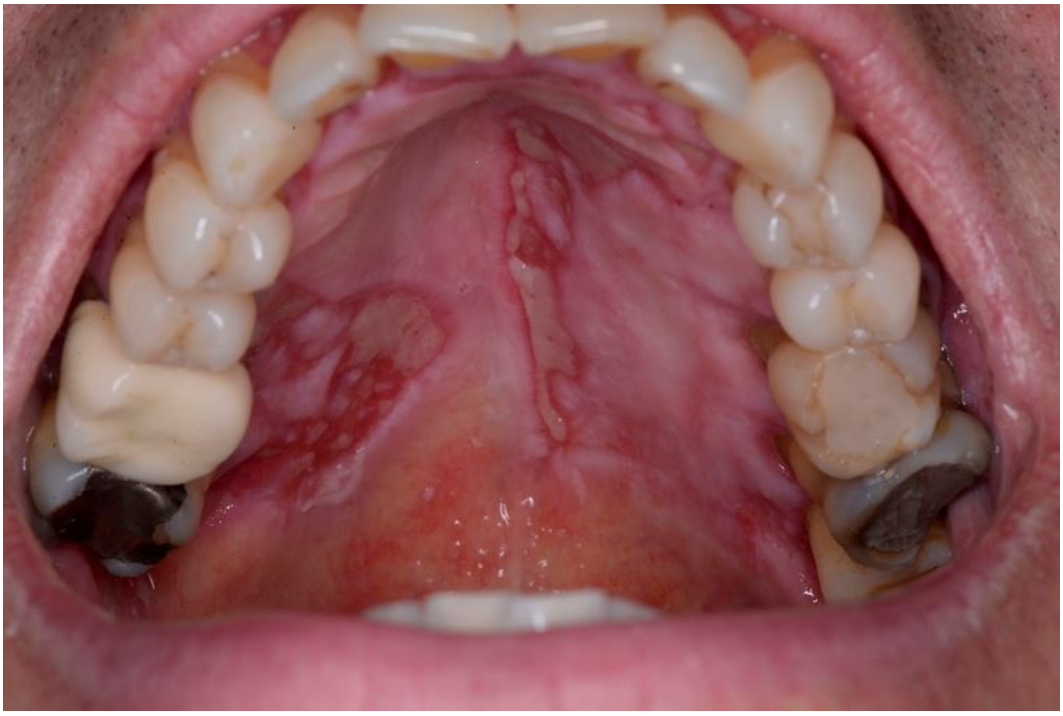
bjd_18566_f1.jpg



bjd_18566_f2a.tiff



bjd_18566_f2b.tiff



bjd_18566_f2c.tiff