**Validity and responsiveness of pain rating scales in patients with chronic oral mucosal diseases.**

Paswach Wiriyakijja1, 2\*, Stephen Porter1, Stefano Fedele1, 3, Tim Hodgson4, Roddy McMillan4, Martina Shephard4, Richeal Ni Riordain1, 5

*1 UCL Eastman Dental Institute, London, United Kingdom*

*2 Department of Oral Medicine, Faculty of Dentistry, Chulalongkorn University, Bangkok, Thailand*

*3 NIHR University College London Hospitals Biomedical Research Centre, London, United Kingdom*

*4 Royal National ENT & Eastman Dental Hospitals, UCLH Foundation NHS Trust London, London, United Kingdom*

*5 Department of Oral Medicine, Cork University Dental School and Hospital, University College Cork, Ireland.*

*\*Corresponding author.*

*E-mail address:* [*paswach.w@gmail.com*](mailto:paswach.w@gmail.com)

**Running title:** Validation of pain scales in oral mucosal diseases

**Key words:** Pain; Validation; Oral lichen planus; Recurrent aphthous stomatitis; Oral mucosal diseases

**ABSTRACT**

**Objectives:** To validate the Visual Analog Scale (VAS) and Numerical Rating Scale (NRS) for measuring pain intensity in chronic oral mucosal diseases.

**Methods:** Secondary analyses of data including the VAS, NRS, demographic, clinical and quality-of-life outcomes at baseline and 4-month follow up were retrieved from a clinical study of chronic oral mucosal diseases. Construct and criterion validity and responsiveness of the VAS and NRS were assessed through testing hypotheses based upon strength of Spearman’s correlation coefficients.

**Results:** Data of 500 and 290 patients with chronic oral mucosal diseases were used for the assessment of validity and responsiveness, respectively. Moderate-to-high correlations between both pain scores and scores of clinical and quality-of-life outcomes were observed, supporting construct validity of the VAS and NRS. Their criterion validity was confirmed by significantly strong association between scores of both scales. Responsiveness of both scales was adequate based on moderate association between their change scores and global rating of change scale.

**Conclusion:** The present results provide evidence supporting validity and responsiveness of the VAS and NRS for pain intensity assessment in patients with chronic oral mucosal diseases. Future research examining other pain intensity domains and standardizing composite scores for pain intensity in this population are required.

**Introduction**

Chronic oral mucosal diseases (COMD) are a diverse group of long-standing conditions that give rise to a wide range of oral mucosal symptoms and signs (Ní Ríordáin & Wiriyakijja, 2017). Some of these diseases are common such as recurrent aphthous stomatitis (RAS) and oral lichen planus (OLP) whereas others including mucous membrane pemphigoid (MMP) and pemphigus vulgaris (PV) are less prevalent. Pain is major subjective symptom experienced by patients with COMD (Ni Riordain et al., 2011), and standardized pain measurement is therefore crucial for baseline assessment before initiating treatment, clinical monitoring and evaluation of the effectiveness of therapeutic interventions in this patient group (Abdalla-Aslan et al., 2016).

There are various facets of pain measurement, which include but not limited to pain intensity, pain frequency, pain quality, temporal features of pain, pain behaviour and pain interference (Chiarotto et al., 2015). Among these pain domains, its sensory intensity is usually the most commonly assessed aspect of clinical pain. Pain intensity refers to strength or ‘loudness’ of pain, and demonstrates ‘how much a patient hurts, reflecting overall magnitude of the pain experience’ (Chiarotto et al., 2019, Fillingim et al., 2016). Different patient reported outcome measures are available for assessing pain intensity ranging from single-item rating scale such as Visual Analog Scale (VAS), Numerical Rating Scale (NRS), Faces Pain Scale and Verbal Rating Scale (VRS; e.g. mild, moderate and severe) to multi-item instruments including Brief Pain Inventory (BPI), Graded Chronic Pain Scale (GCPS) and Multidimensional Pain Inventory (MPI) (Fillingim et al., 2016).

Based upon previous comprehensive reviews, the VAS is the most frequently adopted measure of pain intensity in clinical studies (including RCTs) of COMD (Wiriyakijja et al., 2018, Wiriyakijja et al., 2017, Yan et al., 2020), followed by the NRS. The VAS possesses ratio properties, the highest level of scale measurement hierarchy, allowing for all mathematical operations and analysis on a continuum of pain intensity scores. Despite its dominance in the COMD literature, the VAS has certain limitations including difficulty in completion in the older population, scoring (additional step to measure the length of patient’s rating) and mode of administration that requires written format (either paper or electronic form), and this can be problematic for illiterate or visually impaired individuals (Hawker et al., 2011, Jensen et al., 2001). Alternatively, the NRS, a segmented numeric version of the VAS, has advantages over the VAS in terms of its immediate scoring, scale comprehensibility, and ease of administration (via either written or verbal) (Hawker et al., 2011).

Regarding the validation evidence of the VAS and NRS for measuring pain intensity in COMD, little progress has been achieved in the assessment of measurement properties of both pain scales for use in patients with COMD. One small study found good criterion validity of both instruments for OLP and superior construct validity of the NRS over the VAS based on its greater association with clinical signs of OLP (Chainani-Wu et al., 2008). Importantly, strength of validity evidence of patient-reported scales is established by accumulating validity evidence from results of multiple studies (de Vet et al., 2015), and hence results from a single study cannot provide great confidence in validity of both pain measures in the OLP population. Another UK study found similarity of both pain scales in their responsiveness to changes in OLP disease status over time (Wiriyakijja et al., 2020b). Notably, there remains no studies investigating validity and responsiveness of the VAS and NRS in other groups of COMD including RAS, PV and MMP, questioning the appropriateness of both scales in measuring pain intensity in clinical studies and research of these patient groups.

The primary objective of the present study was to validate the VAS and NRS for measuring pain intensity in a large cohort of patients with COMD by testing construct validity, criterion validity and responsiveness of both instruments for use in this patient group.

**Methods**

***Study design***

The present validation study was a longitudinal secondary analysis of data from the Determination of Minimal Important Difference and Patient Acceptable Symptom State of Patient Reported Outcome Measures in Immunologically mediated Oral Mucosal Diseases (MEAN-IT) study, which was approved by the London – Queen Square Research Ethics Committee (REC reference 17/LO/1825; approval date 3 November 2017). The study approval by the REC was in accordance with the World Medical Association Declaration of Helsinki (Health Research Authority, 2017).

***Participants***

Data for cross-sectional validity testing of the VAS and NRS were obtained from a convenient sample of 500 adult patients with the definitive diagnosis of either OLP, RAS, PV or MMP, who attended their routine review appointments at the oral medicine clinic, UCLH Eastman Dental Hospital, London, United Kingdom from January 2018 to August 2019. Data of a subset of 290 participants in this patient cohort who presented for 4-month follow-up visit were used for longitudinal analysis of responsiveness of both pain scales. The sample size in the present study was in line with recommendation from the COnsensus‐based Standards for the selection of health Measurement INstruments (COSMIN) Risk of Bias checklist, which stipulates that a sample size for a validation study should include at least 100. The eligibility criteria of study participants are shown on Table 1. All potentially eligible participants, in all specialist oral medicine clinics were invited to participate (conducted by PW). All MEAN-IT participants provided verbal and written informed consent to take part in the study.

***Procedures***

A comprehensive oral examination was performed on all study participants. Disease activity of COMD was assessed using the Oral Disease Severity Score (ODSS) in patients with OLP, PV and MMP, and the Ulcer Severity Score (USS) in patients with RAS (all were performed and recorded by PW (D.D.S. (hons), MSc with distinction (Oral Medicine)). Following that, each participant was asked to fill in a set of patient-reported questionnaires including pain scales (VAS, NRS) and the 15-i

tem Chronic Oral Mucosal Disease Questionnaire (COMDQ-15). In the follow-up visit, participants were asked to rate changes in their COMD status on a 7‐point patient global rating of change scale. Demographic and selected clinical data of study participants were obtained from electronic medical records.

***Outcomes***

Primary outcomes of the present study was evidence supporting construct validity, criterion validity and responsiveness of the VAS and NRS for measuring pain intensity for use in patients with COMD.

***Outcome measures***

*Measures of oral pain*

*The Visual Analog Scale (VAS) for pain* is a 100-mm horizontal line with extreme limits of pain intensity including ‘no pain’ and ‘worst pain imaginable’ anchored on the left and right end, respectively. Patients were asked to draw a vertical line on the horizontal scale to indicate a point corresponded to the average intensity of oral pain they were experiencing. The distance from the left end to this vertical mark was recorded in millimetres and denoted the VAS score.

*The Numerical Rating Scale (NRS) for pain* is a 11-point numerical scale with the whole number 0 (no pain) to 10 (worst pain imaginable) presenting in order. Participants were instructed to select one number that was best representative of their pain intensity.

*Measures of disease activity*

The Oral Disease Severity Score (ODSS) is a standardized comprehensive measure of disease activity of oral mucosal diseases, which was first validated for use in OLP patients, and subsequently validated in a group of patients with oral PV and MMP (Escudier et al., 2007, Ormond et al., 2018, Ormond et al., 2020). The ODSS assesses involvement (ODSS-site) and severity (ODSS-activity) of oral mucosal lesions in 17 subsites of the oral cavity. A total ODSS score is the addition of ODSS-site and -activity scores with a score of 11-point verbally administered NRS for average oral pain over the last 2 weeks, with theoretical combined scores ranging from 0 to 106.

*The Ulcer Severity Score (USS)* is the only validated scoring method for quantifying disease activity of RAS, which involves assessment of six RAS-related characteristics including number, size, duration, ulcer-free period, involved sites and pain over the past three months (Tappuni et al., 2013). All RAS-related parameter scores are combined giving total USS score with the maximum total score of 80. For the analysis of the present study, severity of clinical signs of RAS was determined by summation of scores of USS-size and USS-number.

*Measure of quality of life*

The 15-item Chronic Oral Mucosal Disease Questionnaire (COMDQ-15) is a recently developed and validated instrument measuring quality of life specific to patients with COMD (Wiriyakijja et al., 2020a, Wiriyakijja et al., 2020b). Content of the COMDQ-15 is based on inputs from patients with COMD (Ni Riordain et al., 2011), and encompasses 4 relevant quality-of-life domains including Physical Discomfort (PD), Medication and Treatment (MT), Social and Emotional (SE), and Patient Support (PS). Total score of the COMDQ-15 is summation of four subscale scores providing maximum possible score of 60.

*Measure of change*

Global rating of change scale (GRC) was used as reference of change in COMD disease status: “Thinking about all the ways your symptoms related to your oral mucosal conditions are affecting you, compared to the beginning of the study (4 months ago) how do you evaluate the severity of your oral mucosal conditions now?”. The response options are on a 7‐point Likert‐type scale that includes “very much better” (3), “moderately better” (2), “slightly better” (1), “about the same” (0), “slightly worse” (−1), “moderately worse” (−2) and “very much worse” (−3).

***Statistical analysis***

All statistical analyses were performed using STATA version 15.1 (StataCorp). Descriptive analyses of demographics and COMD‐related characteristics were summarised using frequencies and percentages for categorical variables, while mean and standard deviation (sd) were used as summary statistics for continuous variables. Construct validity refers to the extent to which scores of scales of interest validly measures underlying constructs to be measured, and can be assessed through testing hypotheses concerning expected relationship between measuring construct and other related concepts. For the present study, the following hypotheses were postulated:

1. Positive and at least moderate correlations between level of pain intensity (VAS and NRS) and severity of clinical signs of COMD (ODSS-activity for OLP, PV and MMP; USS-size + USS-number for RAS).
2. Positive and at least moderate correlations between level of pain intensity (VAS and NRS) and overall disease activity of COMD (ODSS-total for OLP, PV and MMP; USS-total for RAS)
3. Positive and at least moderate correlations between level of pain intensity (VAS and NRS) and physical discomfort, social and emotional impact, and overall quality of life (COMDQ-15-PD, -SE and -total).

Criterion validity demonstrates the extent to which scores of scales of interest are an adequate reflection of a criterion measure of the same concept of interest. For the present study, criterion validity was assessed by examining whether correlations between the VAS and NRS for pain intensity at both baseline and follow-up visits were positively strong. Evidence supporting responsiveness or sensitivity of a scale to detect change in measuring construct over time was obtained by testing hypotheses that change scores of pain intensity scales and criterion measure of change (GRC) will have moderate and positive correlations. Spearman correlation coefficients (rspearman) were calculated to test all predefined hypotheses, with rspearman values of 0.3 or less, between 0.3 to 0.6, and 0.6 or more indicating low, moderate and high correlations, respectively (Dancey & Reidy, 2007).

**Results**

*Sample description*

Study sample consisted of 500 participants with COMD enrolled in the MEAN-IT study, including 300 OLP patients, 120 RAS patients, 32 PV patients and 48 MMP patients. Each study sample completed all studied variables without missing values. The mean age of study cohort was 58.6 years, and 359 (71.8%) were female. In comparison with other studied conditions, participants with RAS in the present cohort reported the highest average scores of both the VAS and NRS for pain. Details of demographics and clinical characteristics of study sample in total and by studied diseases are summarized in Table 2.

*Construct validity*

Using Spearman’s correlation coefficients, moderate-to-high positive correlations between scores of both pain intensity scales and severity of clinical signs and overall disease activity of the whole cohort of patients with COMD and subgroups of patients by diseases were observed. The strongest and weakest magnitudes of associations were found in the PV and RAS subgroup, respectively. Regarding the correlations between scores of pain intensity and aspects and overall level of quality of life, rspearman values were also in moderate-to-high levels for total sample and COMD subgroups except for correlations between COMDQ-SE and both VAS and NRS in the PV subgroup, which were found to be low. The associations of both pain intensity scales with all studied variables were relatively identical in total COMD cohort, although the magnitudes of correlations were slightly greater in all pairs of associations using the VAS when compared to the NRS (Table 3).

*Criterion validity*

All Spearman’s correlation coefficients between the VAS and NRS for pain showed significantly strong association between both pain intensity measures at all study visits and between disease subgroups (rspearman > 0.9 (total COMD) and > 0.8 (by each disease), P-values < 0.01) as displayed on Table 4.

*Responsiveness*

Significantly positive and moderate associations between GRC and change scores of the VAS and NRS of the total COMD cohort (N = 290) including OLP, RAS and MMP subgroups were noted (Table 5). For the PV subgroup, only the association of VAS change scores and GRC was statistically significant (rspearman = 0.431, P-value < 0.05).

**Discussion**

The present secondary validation analyses demonstrated that both the commonly used VAS and NRS have sufficient construct validity, criterion validity and responsiveness for measuring pain intensity associated with COMD. As for construct validity, all correlation pairs between pain intensity scales and measures of clinical signs, overall disease activity and quality of life were moderate to high, fulfilling the predefined hypotheses in total COMD cohort. The present results also confirmed good level of criterion validity of the two measures based upon a significantly strong correspondence between the scores of VAS and NRS for pain in total COMD sample (rspearman = 0.94-0.97). Moderate positive correlation between change scores of the VAS and NRS and global assessment of change in the whole COMD sample also provide some evidence supporting responsiveness of both pain intensity scale in this patient population.

The main strength of the present study is the relatively large number of sample size of patients with COMD used to validate the VAS and NRS when compared to the only one validation study of both scales in a subgroup of patients with OLP (Chainani-Wu et al., 2008). Although some measurement properties of the VAS and NRS have been previously investigated, these tests were based on less than 50 participants, which were considered inadequate leading to lower quality of supporting psychometric evidence in this patient population. The results of the present study therefore provided additional and more robust evidence supporting the use of both pain intensity scale in the COMD population.

In comparison to the existing literature, the present results supported the findings from a study by Chainani-Wu and colleagues (2007) regarding very high correlation strengths between the VAS and NRS (rPearson = 0.91-0.97) in a smaller cohort of OLP patients in the US. The present findings also showed comparable correlations between both pain intensity scales and other studied variables (including clinical scoring) although the VAS correlated slightly better than the NRS. This was contradictory to the previous psychometric findings, which found stronger correlations of the NRS with clinical signs than those of the VAS. This might be partly due to difference in choices of measures of clinical signs and study populations. Also, the present results might be somewhat explained by superior ability of the VAS in measuring across continuous values of subjective outcomes while the NRS can only yield ordinal scores in whole number.

Regarding the validity and responsiveness of the VAS and NRS based upon each studied condition, the results are relatively consistent among OLP, RAS and MMP subgroups. In contrast, these findings appeared not to be applicable to the PV subgroup according to its low and insignificant correlation results particularly in the testing of responsiveness. This might be due to lower number of PV patients in this cohort, the majority of which were recall patients who had relatively mild pain intensity level and disease severity from the long-term use of systemic corticosteroids and/or immunosuppressive therapies (in over 80% of cases), while most patients in other COMD received only topical regimens, and their oral conditions might not be as controllable as in the PV subgroup. Therefore, the present PV cohort might not be a good representation of the PV population, and future responsiveness studies of the VAS and NRS with larger sample size of PV are recommended.

Apart from relatively low PV sample size, Interpretation of the present results should be cautious due to other study limitations. The present study only validated the VAS and NRS for the measurement of average pain intensity associated with COMD as a single score, and future research using different pain intensity domains (e.g. worst or least pain intensity) or generating composite ratings to improve validity of the scale in assessing treatment effects are recommended. The present validation evidence limits its generalizability as the study participants were recruited from only one tertiary Oral Medicine clinic in the UK, which appeared to be different from real-world population of COMD. Also, data on pain perception can be influenced by confounders including education level and sociocultural factors, and thus cross-cultural validation of both pain intensity scales in population of different sociocultural background could improve generalizability of the present results.

Importantly, the judgement whether which pain intensity scale performs better than another based upon correlation coefficients or psychometric properties (validity, reliability, responsiveness) alone can be misleading (Hjermstad et al., 2011), and to make informed decision on this matter, other important properties of the scale including scale comprehensibility and clinical applicability as well as patient’s compliance and preference should be taken into consideration. A recent comprehensive review of empirical evidence supporting selection of pain scales in adults found that the NRS was often the most recommended single-item pain scale for use in adult patients without cognitive impairment (Safikhani et al., 2018). The NRS was reported to be pain intensity scale of choice in age-mixed population, in chronic pain patients and head-and-neck cancer patients while less educated patients or the elderly preferred the Verbal Rating Scale (e.g. mild, moderate, severe) over the NRS. Importantly, none of previous empirical studies recommended the VAS as preferred scale. Key features of useful outcome measures for routine clinical practice should include simplicity, brevity, quick completion and ease of scoring, and the NRS appeared to comply with these characteristics than the VAS (Renskers et al., 2018).

Apart from psychometric strengths and operational characteristics of outcome measures, there are two additional factors requiring more attention when assessing pain intensity in the clinical trials. One is specific pain intensity domain(s) being assessed, which was found to be reported inconsistently in the literature of chronic pain and COMD including average, worst, or current pain intensity (Wiriyakijja et al., 2018, Jensen et al., 2015, Yan et al., 2020). While a patient’s average pain intensity is arguably the most important outcome domain to target for pain control, a recent study comparing average and worst pain scores found that worst pain ratings were more valid and strongly associated with disabilities than other pain intensity domains (Jensen et al., 2015). Another is the assessment approach to evaluate that specific domain. Although a single rating of recalled average pain over the last 24 hours was the most frequently reported method, a recent finding showed that a composite score combining two recalled pain ratings appeared to be adequate for detecting treatment effects than a single pain rating (Jensen et al., 2015). Obviously, the present study results only provided initial empirical evidence to pave the way towards standardization of pain outcome assessment in COMD, and more studies aiming to resolve important issues related to pain intensity assessment in COMD are recommended.

**Conclusion**

The present study provides some evidence supporting construct validity, criterion validity and responsiveness of the VAS and NRS for the assessment of average level of oral pain in patients with COMD. Both pain intensity unidimensional scales have adequate and relatively comparable measurement properties for use in clinical practice and research of COMD. For standardized pain measurement in COMD, future research examining different pain intensity domains and standardizing a valid composite score for pain intensity measure in COMD are required before inclusion of pain intensity scale as one of a core outcome measure in COMD.

**Acknowledgements**

Paswach Wiriyakijja would like to thank Dr Emma Hayes, Dr Priya Thakrar, Dr Krupali Patel, Dr Barbara Carey, Dr Carolina Venda Nova, Dr Craig Whitelaw, Dr Sanjeet Singhota, Dr Thomas Saunsbury and Dr Valeria Mercadante for their substantial help and supports towards the recruitment process of the study.

**Author contributions**

Paswach Wiriyakijja contributed to conceptualization, methodology, data acquisition, formal analysis and investigation, and drafted the original and revised manuscript. Stephen Porter and Stefano Fedele provided supervision, resource and drafted the original manuscript. Tim Hodgson and Roddy McMillan provided resource and drafted the original manuscript. Martina Shephard provided resource. Richeal Ni Riordain provided supervision and drafted the original and revised manuscript.

**Compliance with Ethical Standards**

**Conflict of interest**

The authors declare that they have no conflicts of interest.

**Funding**

Paswach Wiriyakijja received a PhD Scholarship from the Faculty of Dentistry, Chulalongkorn University, Bangkok, Thailand. Stefano Fedele received funding from the National Institute for Health Research University College London Hospitals Biomedical Research Centre. The MEAN-IT study received service support funding from the National Institute for Health Research University College London Hospitals Biomedical Research Centre and the NIHR Clinical Research Network Portfolio.

**Ethical approval**

All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards (REC reference number: 17/LO/1825).

**Informed consent**

Informed consent was obtained from all individual participants included in the study.

**References**

Abdalla-Aslan R, Benoliel R, Sharav Y and Czerninski R (2016). Characterization of pain originating from oral mucosal lesions. *Oral surgery, oral medicine, oral pathology and oral radiology* **121**: 255-61.

Chainani-Wu N, Silverman S, Jr., Reingold A, Bostrom A, Lozada-Nur F and Weintraub J (2008). Validation of instruments to measure the symptoms and signs of oral lichen planus. *Oral surgery, oral medicine, oral pathology, oral radiology, and endodontics* **105**: 51-8.

Chiarotto A, Deyo RA, Terwee CB, Boers M, Buchbinder R, Corbin TP, Costa LOP, Foster NE, Grotle M, Koes BW, Kovacs FM, Lin C-WC, Maher CG, Pearson AM, Peul WC, Schoene ML, Turk DC, van Tulder MW and Ostelo RW (2015). Core outcome domains for clinical trials in non-specific low back pain. *European Spine Journal* **24**: 1127-1142.

Chiarotto A, Maxwell LJ, Ostelo RW, Boers M, Tugwell P and Terwee CB (2019). Measurement Properties of Visual Analogue Scale, Numeric Rating Scale, and Pain Severity Subscale of the Brief Pain Inventory in Patients With Low Back Pain: A Systematic Review. *The journal of pain : official journal of the American Pain Society* **20**: 245-263.

Dancey CP and Reidy J (2007). *Statistics without maths for psychology,* [s.n.]: [S.l.].

de Vet HCW, Terwee C, Mokkink W and Knol DL (2015). *Measurement in medicine : a practical guide*.

Escudier M, Ahmed N, Shirlaw P, Setterfield J, Tappuni A, Black MM and Challacombe SJ (2007). A scoring system for mucosal disease severity with special reference to oral lichen planus. *The British journal of dermatology* **157**: 765-70.

Fillingim RB, Loeser JD, Baron R and Edwards RR (2016). Assessment of Chronic Pain: Domains, Methods, and Mechanisms. *The journal of pain : official journal of the American Pain Society* **17**: T10-20.

Hawker GA, Mian S, Kendzerska T and French M (2011). Measures of adult pain: Visual Analog Scale for Pain (VAS Pain), Numeric Rating Scale for Pain (NRS Pain), McGill Pain Questionnaire (MPQ), Short-Form McGill Pain Questionnaire (SF-MPQ), Chronic Pain Grade Scale (CPGS), Short Form-36 Bodily Pain Scale (SF-36 BPS), and Measure of Intermittent and Constant Osteoarthritis Pain (ICOAP). *Arthritis care & research* **63 Suppl 11**: S240-52.

Health Research Authority (2017). UK policy framework for health and social care research.

Hjermstad MJ, Fayers PM, Haugen DF, Caraceni A, Hanks GW, Loge JH, Fainsinger R, Aass N and Kaasa S (2011). Studies comparing Numerical Rating Scales, Verbal Rating Scales, and Visual Analogue Scales for assessment of pain intensity in adults: a systematic literature review. *Journal of pain and symptom management* **41**: 1073-93.

Jensen M, Karoly P, Turk D and Melzack RJLG (2001). Handbook of pain assessment.

Jensen MP, Castarlenas E, Tomé-Pires C, de la Vega R, Sánchez-Rodríguez E and Miró J (2015). The Number of Ratings Needed for Valid Pain Assessment in Clinical Trials: Replication and Extension. *Pain medicine (Malden, Mass.)* **16**: 1764-72.

Ni Riordain R, Meaney S and McCreary C (2011). Impact of chronic oral mucosal disease on daily life: preliminary observations from a qualitative study. *Oral diseases* **17**: 265-9.

Ní Ríordáin R and Wiriyakijja P (2017). Patient reported outcome and experience measures of oral disease in oral medicine. *British dental journal* **223**: 713.

Ormond M, McParland H, Donaldson ANA, Andiappan M, Cook RJ, Escudier M, Hullah E, Higham J, McMillan R, Taylor J, Shirlaw PJ, Challacombe SJ and Setterfield JF (2018). An Oral Disease Severity Score validated for use in oral pemphigus vulgaris. *The British journal of dermatology* **179**: 872-881.

Ormond M, McParland H, Thakrar P, Donaldson ANA, Andiappan M, Cook RJ, Escudier ME, Higham J, Hullah E, McMillan R, Taylor J, Shirlaw PJ, Challacombe SJ and Setterfield JF (2020). Validation of an Oral Disease Severity Score (ODSS) tool for use in oral mucous membrane pemphigoid. *The British journal of dermatology* **183**: 78-85.

Renskers L, van Uden RJJC, Huis AMP, Rongen SAA, Teerenstra S and van Riel PLCM (2018). Comparison of the construct validity and reproducibility of four different types of patient-reported outcome measures (PROMs) in patients with rheumatoid arthritis. *Clinical Rheumatology* **37**: 3191-3199.

Safikhani S, Gries KS, Trudeau JJ, Reasner D, Rüdell K, Coons SJ, Bush EN, Hanlon J, Abraham L and Vernon M (2018). Response scale selection in adult pain measures: results from a literature review. *J Patient Rep Outcomes* **2**: 40-40.

Tappuni AR, Kovacevic T, Shirlaw PJ and Challacombe SJ (2013). Clinical assessment of disease severity in recurrent aphthous stomatitis. *Journal of oral pathology & medicine : official publication of the International Association of Oral Pathologists and the American Academy of Oral Pathology* **42**: 635-41.

Wiriyakijja P, Fedele S, Porter S, Mercadante V and Ni Riordain R (2017). Patient-reported outcome measures in recurrent aphthous stomatitis: A critical assessment of quality properties. *Oral diseases* **23**: 1168-1179.

Wiriyakijja P, Fedele S, Porter SR, Mercadante V and Ni Riordain R (2018). Patient-reported outcome measures in oral lichen planus: A comprehensive review of the literature with focus on psychometric properties and interpretability. *Journal of oral pathology & medicine : official publication of the International Association of Oral Pathologists and the American Academy of Oral Pathology* **47**: 228-239.

Wiriyakijja P, Porter S, Fedele S, Hodgson T, McMillan R, Shephard M and Ni Riordain R (2020a). Development and validation of a short version of Chronic Oral Mucosal Disease Questionnaire (COMDQ-15). *Journal of oral pathology & medicine : official publication of the International Association of Oral Pathologists and the American Academy of Oral Pathology* **49**: 55-62.

Wiriyakijja P, Porter S, Fedele S, Hodgson T, McMillan R, Shephard M and Ni Riordain R (2020b). Meaningful improvement thresholds in measures of pain and quality of life in oral lichen planus. *Oral diseases*.

Yan Y-R, Hua F, He M-J, Lei T, Tan Y-Q and Zhou G (2020). Heterogeneity of Outcome Measures Used in Randomized Controlled Trials for the Treatment of Oral Lichen Planus: A Methodological Study. *Journal of Evidence Based Dental Practice* **20**: 101468.

**Table 1** Study eligibility criteria

|  |  |
| --- | --- |
| Inclusion criteria | Exclusion criteria |
| - Aged 18 years or older  - Able to understand and complete  questionnaires | - Having coexisting chronic neuropathic orofacial pain, such as post-traumatic trigeminal neuropathic pain, persistent idiopathic facial pain or burning mouth syndrome |
| - Agree to participate and provide written  informed consent | - Severe systemic disease (ASA 3 or more) and/or some psychiatric conditions which might affect the participation of the study such as schizophrenia |
| Having one of the following conditions |  |
| 1. Oral lichen planus | |
| - Clinically and histopathologically-confirmed  OLP based upon modified WHO  diagnostic criteria | - Evidence of oral epithelial dysplasia in biopsy  specimen  - Evidence of proven hypersensitivity to dental  materials and oral care products  - Evidence of clear temporal relationship of the  development of lesions after the initiation of systemic  medications  - Evidence of oral lichenoid lesions associated with   graft-versus-host disease and systemic lupus  erythematosus |
| 2. recurrent aphthous stomatitis | |
| - Having recurrent oral ulceration (ulcer  episodes of at least twice a year) | - Having RAS-like ulcerations associated with systemic  disorders such as Behçet’s disease, Sweet syndrome,  Ulcerative colitis, Crohn’s disease, Coeliac disease,  auto-inflammatory syndromes, or haematological  abnormalities (severe anaemia, cyclic or chronic  neutropenia) |
| 3. pemphigus vulgaris | |
| - DIF/IIF or ELISA-proven PV |  |
| 4. mucous membrane pemphigoid | |
| - DIF/IIF or ELISA-proven MMP |  |

**Table 2** Demographic and clinical characteristics of study sample (N=500)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Sample characteristics** | OLP  (N=300) | RAS  (N=120) | PV  (N=32) | MMP  (N=48) | P-value | Total  (N=500) |
| Age (y): mean ± sd | 63.2 ± 11.5 | 43.4 ± 13.7 | 59.4 ± 15.9 | 68.1 ± 9.1 | <0.001 | 58.6 ± 15.0 |
| Female: n (%) | 234 (78) | 71 (59.2) | 22 (68.8) | 32 (66.7) | 0.001 | 359 (71.8) |
| Ethnicity: n (%) |  |  |  |  | <0.001 |  |
| White | 204 (68) | 93 (77.5) | 14 (43.8) | 44 (91.7) |  | 355 (71) |
| Mixed | 6 (2) | 5 (4.2) | 2 (6.3) | 0 (0) |  | 13 (2.6) |
| Asian | 79 (26.3) | 16 (13.3) | 14 (43.8) | 3 (6.3) |  | 112 (22.4) |
| Black | 11 (3.7) | 6 (5) | 2 (6.3) | 1 (2.1) |  | 20 (4) |
| Current or previous smokers: n (%) | 72 (24) | 21 (17.5) | 1 (3.1) | 5 (10.4) | 0.007 | 99 (19.8) |
| Alcohol consumers: n (%) | 196 (65.3) | 81 (67.5) | 14 (43.8) | 31 (64.6) | 0.087 | 322 (64.4) |
| Disease duration (y): mean ± sd | 7.9 ± 7.0 | 19.6 ± 15.0 | 3.3 ± 3.7 | 6.0 ± 5.0 | <0.001 | 10.3 ± 10.7 |
| Treatment types: n (%) |  |  |  |  | <0.001 |  |
| Topical treatment | 290 (96.7) | 99 (82.5) | 5 (15.6) | 33 (68.8) |  | 427 (85.4) |
| Systemic ± topical treatment | 10 (3.3) | 21 (17.5) | 27 (84.4) | 15 (31.3) |  | 73 (14.6) |
| Baseline pain: mean ± sd |  |  |  |  |  |  |
| Visual analog scale (0-100mm) | 32 ± 26 | 37 ± 30 | 22 ± 27 | 31 ± 28 | 0.023 | 32 ± 27 |
| Numerical rating scale (0-10) | 3.4 ± 2.5 | 3.9 ± 2.9 | 2.5 ± 2.7 | 3.4 ± 2.6 | 0.047 | 3.4 ± 2.6 |
| NRS = 0 (no pain) | 49 (16.3) | 21 (17.5) | 10 (31.3) | 8 (16.7) | 0.005 | 88 (17.6) |
| NRS = 1-3 (mild pain) | 109 (36.3) | 37 (30.8) | 12 (37.5) | 17 (35.4) |  | 175 (35) |
| NRS = 4-6 (moderate pain) | 103 (34.3) | 27 (22.5) | 7 (21.9) | 15 (31.3) |  | 152 (30.4) |
| NRS = 7-10 (severe pain) | 39 (13) | 35 (29.2) | 3 (9.4) | 8 (16.7) |  | 85 (17) |
| Baseline disease activity: mean ± sd |  |  |  |  |  |  |
| Oral Disease Severity Score (OLP, PV, MMP) | |  |  |  |  |  |
| ODSS-site | 5.9 ± 3.4 | n/a | 4.2 ± 3.3 | 5.5 ± 3.2 | 0.02 | 5.7 ± 3.4 |
| ODSS-activity | 7.1 ± 6.2 | n/a | 7.8 ± 8.1 | 11.3 ± 10.0 | <0.001 | 7.7 ± 7.0 |
| ODSS-total | 16.7 ± 10.7 | n/a | 15.2 ± 13.3 | 20.5 ± 14.4 | 0.063 | 17.0 ± 11.5 |
| Ulcer Severity Score (RAS) |  |  |  |  |  |  |
| USS-size | n/a | 4.5 ± 2.7 | n/a | n/a |  | 4.5 ± 2.7 |
| USS-number | n/a | 3.1 ± 2.3 | n/a | n/a |  | 3.1 ± 2.3 |
| USS-duration | n/a | 3.6 ± 2.1 | n/a | n/a |  | 3.6 ± 2.1 |
| USS-ulcer free | n/a | 7.2 ± 2.6 | n/a | n/a |  | 7.2 ± 2.6 |
| USS-pain | n/a | 5.8 ± 2.3 | n/a | n/a |  | 5.8 ± 2.3 |
| USS-site | n/a | 5.7 ± 2.7 | n/a | n/a |  | 5.7 ± 2.7 |
| USS-total | n/a | 29.9 ± 8.9 | n/a | n/a |  | 29.9 ± 8.9 |
| Baseline quality of life: mean ± sd |  |  |  |  |  |  |
| Chronic Oral Mucosal Disease Questionnaire | |  |  |  |  |  |
| COMDQ-15-Physical Discomfort | 9.6 ± 5.0 | 11.3 ± 4.2 | 9.0 ± 4.7 | 9.6 ± 4.7 | 0.006 | 10.0 ± 4.8 |
| COMDQ-15-Medication & Treatment | 5.8 ± 3.4 | 6.9 ± 2.9 | 7.3 ± 3.6 | 6.0 ± 3.5 | 0.004 | 6.2 ± 3.3 |
| COMDQ-15-Social & Emotional | 6.9 ± 5.1 | 9.0 ± 5.0 | 8.2 ± 5.8 | 5.9 ± 4.7 | <0.001 | 7.4 ± 5.2 |
| COMDQ-15-Patient Support | 2.5 ± 2.2 | 3.2 ± 2.0 | 2.2 ± 1.8 | 2.4 ± 2.1 | 0.01 | 2.6 ± 2.1 |
| COMDQ-15-total | 24.9 ± 12.5 | 30.4 ± 10.4 | 26.8 ± 11.9 | 23.9 ± 11.4 | <0.001 | 26.2 ± 12.1 |

**Table 3** Construct validity: Spearman’s correlation coefficients between scores of pain intensity scales and clinical signs, overall disease activity and quality-of-life outcomes (N = 500)

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Variables | OLP | |  | RAS | |  | PV | |  | MMP | |  | Total | |
| VAS | NRS |  | VAS | NRS |  | VAS | NRS |  | VAS | NRS |  | VAS | NRS |
| *Clinical signs* |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| ODSS-activity | 0.494\*\* | 0.479\*\* |  | n/a | n/a |  | 0.636\*\* | 0.627\*\* |  | 0.320\* | 0.343\* |  | 0.479\*\* | 0.478\*\* |
| USS-size + USS-number | n/a | n/a |  | 0.338\*\* | 0.301\*\* |  | n/a | n/a |  | n/a | n/a |  | 0.338\*\* | 0.301\*\* |
| *Overall disease activity* |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| ODSS-total | 0.648\*\* | 0.635\*\* |  | n/a | n/a |  | 0.753\*\* | 0.743\*\* |  | 0.505\*\* | 0.531\*\* |  | 0.631\*\* | 0.629\*\* |
| USS-total | n/a | n/a |  | 0.483\*\* | 0.440\*\* |  | n/a | n/a |  | n/a | n/a |  | 0.483\*\* | 0.440\*\* |
| *Quality of life* |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| COMDQ-15-PD | 0.653\*\* | 0.649\*\* |  | 0.446\*\* | 0.422\*\* |  | 0.625\*\* | 0.586\*\* |  | 0.539\*\* | 0.547\*\* |  | 0.591\*\* | 0.580\*\* |
| COMDQ-15-SE | 0.529\*\* | 0.533\*\* |  | 0.447\*\* | 0.432\*\* |  | 0.242 | 0.257 |  | 0.622\*\* | 0.617\*\* |  | 0.495\*\* | 0.493\*\* |
| COMDQ-15-total | 0.625\*\* | 0.635\*\* |  | 0.462\*\* | 0.441\*\* |  | 0.436\* | 0.418\* |  | 0.623\*\* | 0.614\*\* |  | 0.568\*\* | 0.567\*\* |

Note: \*P-values < 0.05; \*\*P-values <0.01

**Table 4** Criterion validity: Spearman’s correlation coefficients between scores of the VAS and NRS

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| N | Visits | Spearman's correlation coefficients | | | | |
| OLP | RAS | PV | MMP | total |
| 500 | baseline VAS1-NRS1 | 0.967\*\* | 0.963\*\* | 0.982\*\* | 0.957\*\* | 0.966\*\* |
| 290 | baseline VAS1-NRS1 | 0.963\*\* | 0.965\*\* | 0.937\*\* | 0.921\*\* | 0.964\*\* |
|  | follow-up VAS2-NRS2 | 0.937\*\* | 0.935\*\* | 0.806\*\* | 0.953\*\* | 0.935\*\* |

Note: \*P-values < 0.05; \*\*P-values <0.01

**Table 5** Responsiveness: Spearman’s correlation coefficients between change scores of pain intensity scales and GRC

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Correlations | Spearman’s correlation coefficients | | | | |
| OLP (N=156) | RAS (N=83) | PV (N=24) | MMP (N=27) | total (N=290) |
| VAS change score-GRC | 0.459\*\* | 0.327\*\* | 0.431\* | 0.559\*\* | 0.424\*\* |
| NRS change score-GRC | 0.457\*\* | 0.343\*\* | 0.263 | 0.584\*\* | 0.416\*\* |

Note: \*P-values < 0.05; \*\*P-values <0.01