Meaningful improvement thresholds in measures of pain and quality of life in oral lichen planus

Paswach Wiriyakijja[1, 2]*, Stephen Porter[1], Stefano Fedele[1, 3], Tim Hodgson[4], Roddy McMillan[4], Martina Shephard[4], Richeal Ni Riordain[1, 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] *Eastman Dental Hospital, UCLH Foundation NHS Trust London, London, United Kingdom*

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

**Corresponding author.*

E-mail address: paswach.w@gmail.com

Running title: Meaningful improvement thresholds of QoL in OLP

Key words: Quality of Life; Oral lichen planus; responsiveness; minimal important change; minimal important difference

ABSTRACT

Objectives: To evaluate the responsiveness of measures of pain and oral health-related quality of life (OH-QoL) in patients with oral lichen planus (OLP) and to determine thresholds for minimal important change (MIC) and minimal important difference (MID) for use in this patient population.

Methods: Data from baseline and 4-month follow-up including Visual Analog Scale (VAS), Numerical Rating Scale (NRS), 14-item Oral Health Impact Profile (OHIP-14), 15-item and 26-item Chronic Oral Mucosal Disease Questionnaire (COMDQ-15; COMDQ-26) were collected from 157 patients with OLP. Responsiveness was assessed by testing hypotheses and calculating the area under the curve. MIC and MID were established based on triangulation of distribution-based and anchor-based estimates.

Results: The results supported adequate responsiveness of VAS, NRS, COMDQ-15 and COMDQ-26 for use in OLP while the OHIP-14 demonstrated relatively low sensitivity to detect improvement in the OLP status. Recommended meaningful improvement thresholds were as follows: VAS (MIC 16 mm; MID 18 mm), NRS (MIC/MID 2 points), OHIP-14 (MIC/MID 5 points), COMDQ-15 (MIC 5 points; MID 6 points), and COMDQ-26 (MIC/MID 9 points).

Conclusion: This study provides some evidence of responsiveness as well as establishing meaningful improvement thresholds in scores of pain and OH-QoL measures in OLP.

Introduction

Oral lichen planus (OLP) is a relatively common immune-mediated condition in which patients often experience oral discomfort, reduced oral functioning and significant impairment of quality of life, resulting from persistent inflammation and oral ulceration (Eisen, Carrozzo, Bagan Sebastian, & Thongprasom, 2005). As the disease has no established cure, the primary goal of management of OLP is to lessen oral painful symptoms and improve patients' oral health-related quality of life (OH-QoL) (Thongprasom, Carrozzo, Furness, & Lodi, 2011). Therefore, patient-reported outcomes such as pain and OH-QoL should be used as key outcomes in both clinical practice and studies.

In the last decade the role of patient-reported outcome measures (PROMs) have expanded dramatically, particularly in the assessment of treatment efficacy in the clinical practice and research of chronic medical conditions (FDA, 2009; Kyte, Ives, Draper, & Calvert, 2016). However, the interpretation of the scores generated by PROMs can still be challenging (King, 2011). The scores generated by PROMs to quantify latent (unobservable) constructs such as pain intensity and quality of life may be unfamiliar to both clinicians and researchers (Coon & Cappelleri, 2016). In addition, there may be insufficient available published data to facilitate the interpretation of what, for instance, the magnitude of a 5 point change means on a 0-56 scale of the Oral Health Impact Profile-14 (OHIP-14), or whether a 1-point change in the 0-10 pain scale is clinically relevant to patients.

In research settings, some magnitudes of change can be statistically significant, but this does not necessarily imply clinically important changes (de Vet et al., 2006). This

can happen particularly in the case of studies with large sample sizes that have an increased likelihood of detecting statistical significance when the differences are small (Page, 2014). In order to overcome this issue and to be able to interpret treatment effects, not only does the PROM require to have a good level of responsiveness to detect change in the aspects of a patient's disease status over time, the scores produced by PROMs must also be clinically meaningful (Coon & Cappelleri, 2016; Mokkink et al., 2010). For this purpose, various meaningful change thresholds have been developed to enrich the understanding of the PROM scores, including minimal important change (MIC) and minimal important difference (MID). MIC reflects the smallest magnitude of within-patient change that is clinically important, and is useful to help monitoring patient's disease status in clinical practice (de Vet, Terwee, Mokkink, & Knol, 2015). Whilst MID is the smallest difference in mean scores between groups that could be considered clinically meaningful and is suitable for use in clinical research assessing treatment efficacy (de Vet et al., 2015).

Various measures of pain and OH-QoL have been developed and/or used in clinical practice and research of OLP (Wiriyakijja, Fedele, Porter, Mercadante, & Ni Riordain, 2018). Unfortunately, few studies have evaluated the responsiveness of these instruments (McGrath, Hegarty, Hodgson, & Porter, 2003; Ni Riordain & McCreary, 2012), and surprisingly no studies have examined the clinical meaningfulness of the PROM scores for use in OLP. The primary aim of the present study was to evaluate responsiveness of common measures of pain and OH-QoL as well as establishing meaningful change thresholds including MIC and MID values of these instruments for use in patients with OLP.

Methods

Study design

This was a prospective longitudinal validation study using baseline and 4-month follow-up 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).

Participants

Data were used from a total of 157 patients with OLP who attended regular review appointments at the Oral Medicine clinic, UCLH Eastman Dental Hospital, London, United Kingdom from January 2018 to August 2019. The recruitment of the present study was based upon convenience sampling. All potentially eligible participants in all Consultant lead Oral Medicine clinics were invited to participate. The inclusion and exclusion criteria of study participants are listed in Table 1. After obtaining verbal and written informed consent, all of the participants were prospectively followed from the initial baseline visit to the 4-month follow-up visit.

Sample size

The sample size was in accordance with the recommendation from the consensus-based standards for the selection of health measurement instruments (COSMIN) guideline for systematic reviews of patient-reported outcome measures, which

recommended that a study of responsiveness should include at least 100 subjects to be considered as methodologically sound (Prinsen et al., 2018).

Procedures

After provision of verbal and written informed consent, participants were asked to complete a demographic form (baseline visit only). During both study visits, a comprehensive oral examination was carried out on all study participants to assess oral sites of OLP involvement and disease activity using the Oral Disease Severity Score (Escudier et al., 2007). Participants were categorised into three groups on the basis of the clinical variant of OLP: (i) keratotic (presence of white reticular, papular or plaque-like lesions without apparent erythema/ulceration), (ii) erythematous (presence of atrophic/ erythematous lesions with/without reticular/popular/plaque-like features AND no evidence of erosion/ulceration), and (iii) erosive/ulcerative (presence of erosive or ulcerative lesions with/without the presence of keratotic and/or erythematous changes of OLP) (Bruch & Treister, 2018).

After comprehensive oral examination, participants were then asked to complete a set of patient-reported questionnaires (all listed in the outcome measures section) on both study visits. At the follow-up visit, participants were also asked to respond to an additional question about perception of change in their OLP status on a 7-point patient global rating of change scale. Information regarding medical history, social history and past OLP-related history including disease duration, presence of extra-oral OLP, and current management was obtained from review of electronic patient records.

Outcomes

The outcomes for the primary objective of the present study were as follows: (i) evidence supporting responsiveness to change of the common measures of pain and OH-QoL for use in patients with OLP; (ii) cut-off values corresponding to magnitudes of meaningful change thresholds including the MIC and MID on the scores of the studied measures of pain and OH-QoL.

Outcome measures

Clinical disease activity scoring

The Oral Disease Severity Score (ODSS) is a validated clinical scoring for the measurement of the severity of oral mucosal conditions with special reference to OLP (Escudier et al., 2007). The ODSS assesses the presence, extent and severity of mucosal lesions in 17 oral subsites. A total ODSS score is the addition of clinician-assessed site and activity scores with a score of 0-10 verbal rating scale for average oral pain over the last 2 weeks, with theoretical combined scores ranging from 0 to 106.

Patient-reported outcome measures

The Visual analog scale (VAS) for pain is a measure of pain intensity comprising a 100-mm horizontal line, labeled with ‘no pain’ at one end and ‘worst pain imaginable’ on the other end. Participants were asked to place a vertical mark on the point of the VAS line that best reflected the degree of pain they were currently experiencing from OLP (Hawker, Mian, Kendzerska, & French, 2011).

The Numerical Rating Scale (NRS) for pain estimated severity of oral pain currently experienced by a patient on a whole number scale of 0-10 (11-point scale). Both the

VAS and NRS was validated for use in the OLP population with psychometric evidence supporting their validity and reliability (Chainani-Wu et al., 2008).

The 14-item Oral Health Impact Profile (OHIP-14) is a 14-item, 5-point (0-4) Likert-type questionnaire measuring general OH-QoL on seven domains (each with 2 items) including functional limitation, physical pain, psychological discomfort, physical disability, psychological disability, social disability and handicap. The maximum possible subscale and total score of this scale are 8 and 56, respectively. The greater the OHIP-14 score the poorer of the patient's perception is of their OH-QoL (Slade, 1997).

The 26-item Chronic Oral Mucosal Disease Questionnaire (COMDQ-26) is a 26-item, 5 point (0-4) Likert-type instrument measuring the impact of chronic oral mucosal conditions and the related treatment on a patient's OH-QoL in four different aspects including Pain and Functional limitation (PF, 9 items), Medication and treatment (MT, 6 items), Social and Emotional (SE, 7 items) and Patient Support (PS, 4 items). The total COMDQ-26 scores range from 0 to 104, with the higher scores indicating worse impact of the disease on the patient's OH-QoL (Ni Riordain & McCreary, 2011). The validity and reliability of the COMDQ-26 have been proven acceptable for use in patients with OLP in one study of OLP patients residing in the UK (Ni Riordain, Hodgson, Porter, & Fedele, 2016).

The 15-item Chronic Oral Mucosal Disease Questionnaire (COMDQ-15) is a recently developed short version of the original COMDQ-26 (Wiriyakijja et al., 2020). Similar to its parent version, the COMDQ-15 assesses four OH-QoL domains including Physical Discomfort (PD, 5 items), Medication and Treatment (MT, 3 items), Social and Emotional (SE, 5 items) and Patient Support (PS, 2 items). Total COMDQ-15 score

are calculated by summation of the responses of all items, giving the possible maximum score of 60. The COMDQ-15 has good evidence supporting its validity and reliability for use in patients with OLP (Wiriyakijja et al., 2020).

Anchor question

To assess the responsiveness and meaningful change thresholds of PROMs, criteria are required to confirm whether patients have experienced a change in their disease status - including being worse, improved or stable over time. In this study, the following *patient's global rating of change* (GRC) was used as external anchor/reference of change: "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), 'very much worse' (-3). Participants answering 'moderately better' and 'very much better' were classified as having clinically important improvement, while those responding to the remaining options were considered "not importantly improved".

Statistical analyses

Statistical analyses were performed using STATA version 15.1 (StataCorp, College Station, TX, U.S.A.). Descriptive analyses of demographics and OLP-related characteristics were summarized using frequencies and accompanying percentages for categorical variables, while median and interquartile range (IQR) were used as summary statistics for continuous variables. Score distribution of the studied PROMs

including baseline, follow-up and change scores were presented using mean and standard deviation (SD) based upon the GRC. According to the small sample size of those reporting “very much worse”, “moderately worse” and “slightly worse”, the data were combined and presented as a “worsened” group (n=19). In addition, due to a small sample size in the total “worsened” group, assessment of the responsiveness and meaningful change thresholds were carried out only for the direction of improvement.

Responsiveness

Responsiveness is the ability of PROMs to detect change over time in the construct being measured. Two different approaches were performed to assess responsiveness of the studied PROMs including construct and criterion approaches. For the construct approach, Spearman’s correlation coefficient (rho) was used to test hypotheses of change values of the studied PROM scores and the GRC score. The following hypotheses were formulated:

1. Moderate and positive correlations between GRC scores and change scores of the pain-VAS, pain-NRS, total OHIP-14, total and subscales of the COMDQ-15 and COMDQ-26 (except for the patient support subscale of the COMDQ-15 and COMDQ-26).
2. Low and positive correlations between GRC scores and change scores of the patient support subscale of the COMDQ-15 and COMDQ-26.

Correlation coefficients of 0.3 or less, between 0.3 and 0.6, and 0.6 or greater were defined as low, moderate and high, respectively.

For the criterion approach, responsiveness of the PROMs was examined by checking the area under the curve (AUC) of the Receiver Operating Characteristic (ROC) curve analyses. The AUC represents the ability of PROM scores to correctly identify patients as improved or non-improved based upon the external anchor (GRC). The AUC values of 0.7 or above is considered acceptable (Terwee et al., 2007).

Meaningful improvement thresholds

Two methods were applied for the estimation of meaningful improvement thresholds including distribution-based and anchor-based methods. The distribution-based methods are based solely upon the distributional characteristics of the scores in the sample without the use of external reference. In this study, half a standard deviation at baseline (0.5 $SD_{baseline}$) and standard error of measurement (SEM) were calculated. The SEM was estimated by the following formula: $SEM=SD_{difference}/ \sqrt{2}$, when $SD_{difference}$ is the standard deviation of the difference in scores at baseline and follow-up visit in the group reporting “about the same”.

To determine meaningful within-patient improvement thresholds, anchor-based MIC values were estimated as the ROC cut-off point of change scores of the PROMs with the least amount of misclassified patients between those who were “importantly improved” and “not importantly improved”. In other words, the MIC values were the optimal cut-off points, which maximise true-positive rate (TP; sensitivity) and true-negative rate (TN; specificity) on the ROC curve. To determine meaningful between-group difference thresholds, anchor-based MID values were estimated by calculating the difference in mean change scores of the ‘moderately improved’ and ‘about the same’ group.

Multiple meaningful improvement threshold values from both distribution-based and anchor-based methods were then triangulated to create the recommended thresholds of MIC and MID for each studied PROM score. The triangulation process was based upon average values amongst all estimates with consideration of the limitation of the scale response. For instance, the recommended threshold values were narrowed down to integer value only.

Results

Descriptive statistics

Descriptive summary of baseline demographics and OLP-related characteristics of 157 study participants are present in Table 2. Mean and standard deviation of baseline, follow-up and change scores of all studied PROMs based upon the GRC are shown in Table 3. Of the 157 patients with OLP, 19 (12.1%) reported deterioration [one (0.01%) very much worse, five (0.03%) moderately worse and 13 (0.08%) slightly worse], 52 (33.1%) reported about the same and 86 (54.8%) reported improvement on the GRC.

Responsiveness

For construct approach, predefined hypotheses regarding expected magnitude and direction of correlation between PROM change scores and the GRC, values of Spearman rho coefficients and ascertainment of hypotheses are present in Table 4. The VAS and NRS for pain were similarly moderately responsive to change in OLP disease status over time. The total OHIP-14 was relatively less sensitive to detect

patient's perception of change in OLP status over time compared to the total COMDQ-15 and COMDQ-26. With respect to the COMDQ subscale scores, values of Spearman rho coefficients confirmed the hypotheses in the majority of the subscales except for the MT subscale of the COMDQ-15, which was marginally insufficient to meet the requirement of the hypothesis.

For the criterion approach, the AUC values of change scores of the studied PROMs are present in Table 5. The results showed that only the AUC values of total COMDQ-15 and COMDQ-26, the PF subscale of the COMDQ-26 and the PD subscale of the COMDQ-15 achieved acceptable threshold of responsiveness (0.70).

Meaningful change thresholds

The MIC and MID estimation of all studied PROMs based on distribution-based and anchor-based methods are present in Table 5.

Discussion

The present study examined two important characteristics – responsiveness and interpretability – of common measures of oral symptoms and OH-QoL to support their usage in clinical practice and OLP research. Regarding responsiveness of studied PROMs assessing pain, the present results demonstrated that responsiveness of the VAS and NRS in measuring improvement in patient's perception of OLP status were similar based upon hypothesis testing approach. This is in accordance with one previous study (Chainani-Wu et al., 2008), which found moderate-to-high correlation between the Change in Symptom Scale (CSS) and both measures of oral pain (r_{VAS} =

0.492, r_{NRS} = 0. 549). Based upon the criterion approach, a slightly greater AUC value of the change in the NRS compared to the VAS provides evidence supporting higher accuracy of the former instrument in the detection of change in patients' OLP status over the latter. Considering evidence of responsiveness of the VAS and NRS from both methods, it appears that the NRS is slightly superior to the VAS in its ability to detect improvement in the patient's perception of the OLP status.

As for the responsiveness of the OH-QoL PROMs, the COMDQ-26 was found to be the most sensitive OH-QoL instrument to detect improvement in OLP disease status, followed by the COMDQ-15 and the OHIP-14. Using the generally accepted criteria (AUC of at least 0.70), the present results confirmed adequate evidence supporting the responsiveness to improvement of total COMDQ-15 and COMDQ-26 scores. Regarding the subscale COMDQ scores, PF subscale of the COMDQ-26 and PD subscale of the COMDQ-15 were shown to have acceptable responsiveness to change, while the remaining subscales performed lower than predefined threshold. Considering all of the evidence supporting the responsiveness of the COMDQ, it is recommended to use the total scale scores of both versions over the use of subscale scores, for the assessment of treatment efficacy in OLP.

In comparison, the OHIP-14 showed a poorer level of responsiveness than both the COMDQ versions and all of the included pain scales. One explanation for this finding may be because the OHIP-14 was first developed and validated for use as a self-reported measure of general impact of oral conditions, and mainly for those with dental problems (Slade, 1997). The content of some items of the OHIP-14 such as "have you had painful aching in your mouth?", which appeared to reflect odontogenic pain, rather

than pain associated with oral mucosal conditions, may not always be sensitive enough to detect OLP-related changes. For the continued use of the OHIP-14 in OLP, it is important that researchers or clinicians are aware of the limited content validity and responsiveness of this scale for use in such patients, and further refinement of this widely adopted instrument is therefore required.

To enhance their clinical utility, meaningful improvement thresholds of common measures of pain and OH-QoL were calculated. For research purposes, understanding magnitude of minimal important difference (MID) can be valuable in study designs (e.g. facilitating sample size calculation in studies assessing patient-reported outcomes) as well as assessing treatment efficacy between treatment groups beyond statistical significance (de Vet et al., 2006; Wyrwich, Norquist, Lenderking, & Acaster, 2013). In comparison, the values of minimal important change (MIC) could aid in shared clinical decision-making in the routine clinical setting. For example, It can inform patients and clinicians about the magnitude of change in PROM scores that may justify a change in management, such as introduction of a new treatment, continuation or withdrawal of a current medication, or to increase or decrease the dosage (King, 2011). It can be implied that patients who achieve a score of equal to or greater than thresholds of MIC after a period of treatment may be benefiting from the given intervention.

To the best of our knowledge, this is the first study which has attempted to determine the MID and MIC values for improvement in common measures of pain and OH-QoL in a cohort of patients with OLP. Our results revealed some variability in the values of meaningful improvement thresholds amongst the different quantitative techniques

used. However, the present study adopted a triangulation process, which has been recently recommended by a group of authors, to establish recommended thresholds for further references (Coon & Cappelleri, 2016). It was often observed that the magnitude of within-patient change (MIC) was generally greater than that of between-group difference (MID) (Sully et al., 2019). The present results, however, showed that the values of MIC and MID of studied measures are relatively comparable.

However, it is acknowledged that the present study has several limitations. Due to small sample size of patients whose OLP condition were worsened, only MIC and MID values for improvement were calculated, and these values do not apply for use as reference values for those having a deterioration of the condition. Based on the present results, assessment of responsiveness and meaningful change thresholds for worsening of all studied measures are indeterminate, and future research with larger sample size is recommended. Again due to the small sample size, the present study did not take into consideration the impact of baseline scores, which has been reported to influence the MIC and MID values (Crosby, Kolotkin, & Williams, 2003; Escobar & Riddle, 2014). Regarding generalisability of the present finding, the study cohort in this study was based upon patients in one tertiary referral oral medicine centre, and thus may not represent the real-world OLP population, including asymptomatic cases of OLP. The exclusion of non-English speakers may also reduce the external validity of the study.

In conclusion, the present study provides some evidence of responsiveness to improvement in the VAS, NRS, COMDQ-15 and COMDQ-26 as well as establishing meaningful improvement thresholds of the scores of these instruments. Published

estimates of MID and MIC will allow researchers and clinicians to adopt these as standard for interpretation of improvement of pain and OH-QoL outcomes in OLP.

Author contributions

Paswach Wiriyakijja and Richeal Ni Riordain designed the study. Paswach Wiriyakijja collected data from patients, carried out the statistical analyses reported in the study and drafted manuscript. Richeal Ni Riordain, Roddy McMillan, Martina Shephard, Tim Hodgson, Stefano Fedele and Stephen Porter edited and contributed comments on the manuscripts.

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. 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.

Disclosure statement

No potential conflict of interest was reported by the authors.

References

Bruch, J. M., & Treister, N. S. (2018). *CLINICAL ORAL MEDICINE AND PATHOLOGY*. [Place of publication not identified]: SPRINGER INTERNATIONAL PU.

Chainani-Wu, N., Silverman, S., Jr., Reingold, A., Bostrom, A., Lozada-Nur, F., & Weintraub, J. (2008). Validation of instruments to measure the symptoms and signs of oral lichen planus. *Oral Surg Oral Med Oral Pathol Oral Radiol Endod, 105*(1), 51-58. doi:10.1016/j.tripleo.2007.06.022

Coon, C. D., & Cappelleri, J. C. (2016). Interpreting Change in Scores on Patient-Reported Outcome Instruments. *Ther Innov Regul Sci, 50*(1), 22-29. doi:10.1177/2168479015622667

Crosby, R. D., Kolotkin, R. L., & Williams, G. R. (2003). Defining clinically meaningful change in health-related quality of life. *J Clin Epidemiol, 56*(5), 395-407. doi:10.1016/s0895-4356(03)00044-1

de Vet, H. C., Terwee, C., Mokkink, W., & Knol, D. L. (2015). *Measurement in medicine : a practical guide*.

de Vet, H. C., Terwee, C. B., Ostelo, R. W., Beckerman, H., Knol, D. L., & Bouter, L. M. (2006). Minimal changes in health status questionnaires: distinction between minimally detectable change and minimally important change. *Health Qual Life Outcomes, 4*, 54. doi:10.1186/1477-7525-4-54

Eisen, D., Carrozzo, M., Bagan Sebastian, J. V., & Thongprasom, K. (2005). Number V Oral lichen planus: clinical features and management. *Oral Dis, 11*(6), 338-349. doi:10.1111/j.1601-0825.2005.01142.x

Escobar, A., & Riddle, D. L. (2014). Concordance between important change and acceptable symptom state following knee arthroplasty: the role of baseline scores. *Osteoarthritis Cartilage, 22*(8), 1107-1110. doi:10.1016/j.joca.2014.06.006

Escudier, M., Ahmed, N., Shirlaw, P., Setterfield, J., Tappuni, A., Black, M. M., & Challacombe, S. J. (2007). A scoring system for mucosal disease severity with special reference to oral lichen planus. *Br J Dermatol, 157*(4), 765-770. doi:10.1111/j.1365-2133.2007.08106.x

FDA. (2009). Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims. Guidance for Industry. Retrieved from https://www.fda.gov/regulatory-information/search-fda-guidance-documents/patient-reported-outcome-measures-use-medical-product-development-support-labeling-claims

Hawker, G. A., Mian, S., Kendzerska, T., & 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 Res (Hoboken), 63 Suppl 11*, S240-252. doi:10.1002/acr.20543

King, M. T. (2011). A point of minimal important difference (MID): a critique of terminology and methods. *Expert Rev Pharmacoecon Outcomes Res, 11*(2), 171-184. doi:10.1586/erp.11.9

Kyte, D., Ives, J., Draper, H., & Calvert, M. (2016). Current practices in patient-reported outcome (PRO) data collection in clinical trials: a cross-sectional survey of UK trial staff and management. *BMJ Open, 6*(10), e012281. doi:10.1136/bmjopen-2016-012281

McGrath, C., Hegarty, A. M., Hodgson, T. A., & Porter, S. R. (2003). Patient-centred outcome measures for oral mucosal disease are sensitive to treatment. *Int J Oral Maxillofac Surg, 32*(3), 334-336. doi:10.1054/ijom.2002.0377

Mokkink, L. B., Terwee, C. B., Patrick, D. L., Alonso, J., Stratford, P. W., Knol, D. L., . . . de Vet, H. C. (2010). The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *J Clin Epidemiol, 63*(7), 737-745. doi:10.1016/j.jclinepi.2010.02.006

Ni Riordain, R., Hodgson, T., Porter, S., & Fedele, S. (2016). Validity and reliability of the Chronic Oral Mucosal Diseases Questionnaire in a UK population. *J Oral Pathol Med, 45*(8), 613-616. doi:10.1111/jop.12425

Ni Riordain, R., & McCreary, C. (2011). Validity and reliability of a newly developed quality of life questionnaire for patients with chronic oral mucosal diseases. *J Oral Pathol Med, 40*(8), 604-609. doi:10.1111/j.1600-0714.2011.01021.x

Ni Riordain, R., & McCreary, C. (2012). Further reliability and responsiveness of the Chronic Oral Mucosal Diseases Questionnaire. *Oral Dis, 18*(1), 60-66. doi:10.1111/j.1601-0825.2011.01844.x

Page, P. (2014). Beyond statistical significance: clinical interpretation of rehabilitation research literature. *Int J Sports Phys Ther, 9*(5), 726-736.

Prinsen, C. A. C., Mokkink, L. B., Bouter, L. M., Alonso, J., Patrick, D. L., de Vet, H. C. W., & Terwee, C. B. (2018). COSMIN guideline for systematic reviews of patient-reported outcome measures. *Qual Life Res, 27*(5), 1147-1157. doi:10.1007/s11136-018-1798-3

Slade, G. D. (1997). Derivation and validation of a short-form oral health impact profile. *Community Dent Oral Epidemiol, 25*(4), 284-290. doi:10.1111/j.1600-0528.1997.tb00941.x

Sully, K., Trigg, A., Bonner, N., Moreno-Koehler, A., Trennery, C., Shah, N., . . . Cocks, K. (2019). Estimation of minimally important differences and responder definitions for EORTC QLQ-MY20 scores in multiple myeloma patients. *Eur J Haematol, 103*(5), 500-509. doi:10.1111/ejh.13316

Terwee, C. B., Bot, S. D., de Boer, M. R., van der Windt, D. A., Knol, D. L., Dekker, J., . . . de Vet, H. C. (2007). Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol, 60*(1), 34-42. doi:10.1016/j.jclinepi.2006.03.012

Thongprasom, K., Carrozzo, M., Furness, S., & Lodi, G. (2011). Interventions for treating oral lichen planus. *Cochrane Database Syst Rev*(7), Cd001168. doi:10.1002/14651858.CD001168.pub2

Wiriyakijja, P., Fedele, S., Porter, S. R., Mercadante, V., & 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. *J Oral Pathol Med, 47*(3), 228-239. doi:10.1111/jop.12604

Wiriyakijja, P., Porter, S., Fedele, S., Hodgson, T., McMillan, R., Shephard, M., & Ni Riordain, R. (2020). Development and validation of a short version of Chronic Oral Mucosal Disease Questionnaire (COMDQ-15). *J Oral Pathol Med, 49*(1), 55-62. doi:10.1111/jop.12964

Wyrwich, K. W., Norquist, J. M., Lenderking, W. R., & Acaster, S. (2013). Methods for interpreting change over time in patient-reported outcome measures. *Qual Life Res, 22*(3), 475-483. doi:10.1007/s11136-012-0175-x

Table 1 Study eligibility criteria

Inclusion criteria	Exclusion criteria
- Aged 18 years or older - Clinically and histopathologically proven OLP based upon modified WHO diagnostic criteria (van der Meij & van der Waal, 2003) - Able to understand and complete questionnaires - Agree to participate and provide written informed consent	- Evidence of oral epithelial dysplasia in biopsy specimen - Evidence of proven hypersensitivity to dental materials - Evidence of oral lichenoid lesions associated with graft-versus-host disease and systemic lupus erythematosus - Having coexisting chronic neuropathic orofacial pain, such as post-traumatic trigeminal neuropathic pain, persistent idiopathic facial pain or burning mouth syndrome - Severe systemic disease (ASA 3 or more) and/or some psychiatric conditions which might affect the participation of the study such as schizophrenia

Table 2 Demographic and clinical characteristics of the study sample

Patient characteristics (n=157)	
Demographic variables	
Age (y; median, IQR)	65.5 (55.2, 70.4)
Female (n, %)	122 (77.7)
Ethnicity (n, %)	
White	105 (66.9)
Mixed	5 (3.2)
Asian	40 (25.5)
Black	7 (4.5)
Smoking (n, %)	
Non-smoker	119 (75.8)
Ex-smoker	30 (19.1)
Current smoker	8 (5.1)
Alcohol consumption (n, %)	
No	53 (33.8)
≤ 14 Units/week	89 (56.7)
> 14 Units/week	15 (9.6)
Comorbidity (n, %)	
No	20 (12.7)
1 comorbidity	37 (23.6)
≥ 2 comobidities	100 (63.7)
OLP-related characteristics	
Disease duration (y; median, IQR)	5.5 (2.4, 10.4)
Clinical types	
Keratotic	21 (13.4)
Erythematous	110 (70.1)
Ulcerative	26 (16.6)
Baseline ODSS score (median, IQR)	20 (13, 26)
Baseline ODSS-site	7 (4, 9)
Baseline ODSS-activity	8 (4, 13)
Presence of extraoral LP (n,%)	
Yes	40 (25.5)
Yes/genital	23 (14.7)
Yes/skin	23 (14.7)
Treatment	
Topical Benzydamine	12 (7.6)
TCS	101 (64.3)
TCS + Topical Benzydamine/Lidocaine gel	30 (19.1)
Topical tacrolimus	2 (1.3)
Topical tacrolimus (+/- Topical Benzydamine/TCS)	4 (2.5)
Systemic Prednisolone (+/- Topical Benzydamine/TCS)	2 (1.3)
Systemic Hydroxychloroquine (+/- Topical Benzydamine/TCS)	3 (1.9)
Systemic MMF (+/- Topical Benzydamine/TCS)	2 (1.3)
Systemic AZA (+/- Topical Benzydamine/TCS)	1 (0.6)

Abbreviation: LP = lichen planus; TCS = Topical corticosteroids included at least one or a combination of Betamethasone 0.5 mg tablet as mouthwash, Mometasone furoate 0.1% ointment, Fluticasone Propionate 0.05% spray, Fluticasone propionate 400 µg nasule as mouthwash, Clobetasal propionate 0.05% ointment; MMF = Mycophenolate mofetil; AZA = Azathioprine

Table 3 Descriptive statistics of baseline, follow-up, and change scores of studied PROMs by response categories of the global rating of change scale

Instruments	Baseline scores (mean ± sd)	Follow-up scores (mean ± sd)	Change scores (mean ± sd)
VAS (0-100)			
worse[a] (n=19)	35.1 ± 23.6	57.0 ± 20.6	-21.9 ± 25.6
no change (n=52)	31.8 ± 23.6	32.8 ± 28.0	-0.9 ± 16.8
slightly better (n=37)	44.5 ± 24.8	34.7 ± 22.1	9.8 ± 17.8
moderately better (n=24)	48.2 ± 28.7	19.4 ± 23.3	28.8 ± 24.9
very much better (n=25)	19.7 ± 21.3	8.7 ± 9.0	11.1 ± 20.9
NRS (0-10)			
worse[a] (n=19)	3.4 ± 2.1	5.3 ± 2.4	-1.9 ± 2.2
no change (n=52)	3.5 ± 2.3	3.5 ± 3.0	-0.1 ± 1.7
slightly better (n=37)	4.5 ± 2.3	3.8 ± 2.1	0.7 ± 1.5
moderately better (n=24)	4.9 ± 2.8	2.2 ± 2.1	2.7 ± 1.8
very much better (n=25)	2.3 ± 2.1	1.2 ± 1.2	1.1 ± 2.1
OHIP-14 total (0-56)			
worse[a] (n=19)	23.0 ± 10.7	25.1 ± 11.8	-2.1 ± 8.1
no change (n=52)	19.2 ± 13.0	17.8 ± 13.9	1.3 ± 5.9
slightly better (n=37)	20.6 ± 12.8	18.1 ± 11.4	2.5 ± 6.7
moderately better (n=24)	22.8 ± 14.2	18.2 ± 12.7	4.5 ± 5.1
very much better (n=25)	13.6 ± 11.5	8.0 ± 8.0	5.6 ± 7.2
COMDQ-15 total (0-60)			
worse[a] (n=19)	26.8 ± 10.6	31.7 ± 9.8	-4.8 ± 6.6
no change (n=52)	23.4 ± 11.5	23.3 ± 12.8	0.1 ± 5.7
slightly better (n=37)	26.1 ± 11.1	25.0 ± 11.6	1.1 ± 5.9
moderately better (n=24)	31.8 ± 12.6	25.1 ± 11.0	6.6 ± 7.4
very much better (n=25)	20.3 ± 11.6	13.3 ± 7.2	7.0 ± 9.0
COMDQ-15-PD (0-20)			
worse[a] (n=19)	11.4 ± 4.0	13.1 ± 3.7	-1.6 ± 3.1
no change (n=52)	10.0 ± 5.0	9.6 ± 5.3	0.4 ± 2.3
slightly better (n=37)	10.6 ± 3.8	10.1 ± 4.4	0.6 ± 2.9
moderately better (n=24)	13.1 ± 4.7	10.2 ± 4.7	3.0 ± 3.5
very much better (n=25)	8.6 ± 5.2	5.6 ± 3.2	3.0 ± 4.0
COMDQ-15-MT (0-12)			
worse[a] (n=19)	3.8 ± 3.3	5.3 ± 3.4	-1.5 ± 3.0
no change (n=52)	3.6 ± 3.0	4.0 ± 3.0	-0.4 ± 1.8
slightly better (n=37)	4.8 ± 3.1	4.6 ± 2.8	0.2 ± 1.6
moderately better (n=24)	6.2 ± 3.3	4.9 ± 2.8	1.3 ± 2.0
very much better (n=25)	3.6 ± 3.2	3.0 ± 2.4	0.6 ± 2.6
COMDQ-15-SE (0-20)			
worse[a] (n=19)	8.5 ± 4.6	9.9 ± 4.7	-1.4 ± 2.8
no change (n=52)	6.9 ± 5.1	7.1 ± 5.3	-0.2 ± 3.4
slightly better (n=37)	8.2 ± 5.3	7.4 ± 5.1	0.8 ± 2.9
moderately better (n=24)	9.2 ± 5.8	6.7 ± 4.4	2.5 ± 3.9
very much better (n=25)	6.1 ± 4.4	3.2 ± 3.5	2.9 ± 3.6

Note: [a] worse group (n =19) included 13 slightly worse, 5 moderately worse and 1 very much worse

Table 3 Descriptive statistics of baseline, follow-up, and change scores of studied PROMs by response categories of the global rating of change scale (cont.)

Instruments	Baseline scores (mean ± sd)	Follow-up scores (mean ± sd)	Change scores (mean ± sd)
COMDQ-15-PS (0-8)			
worse[a] (n=19)	3.1 ± 2.2	3.4 ± 1.8	-0.3 ± 1.6
no change (n=52)	2.9 ± 2.3	2.7 ± 2.4	0.2 ± 1.7
slightly better (n=37)	2.6 ± 2.0	3.0 ± 1.8	-0.4 ± 1.7
moderately better (n=24)	3.3 ± 2.8	3.4 ± 2.4	-0.1 ± 2.0
very much better (n=25)	2.1 ± 1.8	1.5 ± 1.8	0.6 ± 1.6
COMDQ-26 total (0-104)			
worse[a] (n=19)	46.1 ± 17.9	53.6 ± 16.9	-7.5 ± 10.0
no change (n=52)	39.7 ± 18.3	39.4 ± 20.9	0.3 ± 9.2
slightly better (n=37)	44.5 ± 17.4	41.9 ± 17.8	2.6 ± 8.7
moderately better (n=24)	52.4 ± 19.4	41.8 ± 16.9	10.6 ± 10.7
very much better (n=25)	35.0 ± 18.2	24.1 ± 12.8	10.9 ± 14.2
COMDQ-26-PF (0-36)			
worse[a] (n=19)	18.2 ± 7.1	20.5 ± 6.9	-2.3 ± 4.1
no change (n=52)	15.5 ± 8.3	14.7 ± 8.7	0.9 ± 4.0
slightly better (n=37)	16.6 ± 6.0	15.5 ± 6.8	1.1 ± 4.0
moderately better (n=24)	20.0 ± 7.8	15.3 ± 7.6	4.7 ± 6.0
very much better (n=25)	13.3 ± 8.0	8.9 ± 6.0	4.4 ± 6.3
COMDQ-26-MT (0-24)			
worse[a] (n=19)	9.7 ± 4.6	12.6 ± 4.5	-2.8 ± 3.6
no change (n=52)	9.1 ± 4.7	9.7 ± 5.1	-0.6 ± 3.2
slightly better (n=37)	10.6 ± 4.9	10.4 ± 4.7	0.3 ± 2.6
moderately better (n=24)	13.3 ± 5.0	10.8 ± 4.3	2.5 ± 3.2
very much better (n=25)	8.5 ± 5.2	7.2 ± 4.2	1.3 ± 4.3
COMDQ-26-SE (0-28)			
worse[a] (n=19)	12.7 ± 6.7	14.6 ± 6.6	-1.9 ± 4.2
no change (n=52)	10.3 ± 7.1	10.6 ± 7.2	-0.2 ± 4.5
slightly better (n=37)	12.3 ± 7.2	11.2 ± 6.9	1.1 ± 4.1
moderately better (n=24)	13.7 ± 7.9	10.1 ± 6.1	3.6 ± 4.6
very much better (n=25)	9.3 ± 6.2	5.2 ± 4.9	4.1 ± 5.1
COMDQ-26-PS (0-16)			
worse[a] (n=19)	5.4 ± 2.8	5.9 ± 3.2	-0.5 ± 1.8
no change (n=52)	4.7 ± 3.2	4.7 ± 3.4	0.1 ± 2.3
slightly better (n=37)	4.9 ± 3.0	5.3 ± 2.8	-0.3 ± 2.4
moderately better (n=24)	5.5 ± 3.7	5.6 ± 2.9	0.0 ± 2.5
very much better (n=25)	3.9 ± 2.0	2.8 ± 2.4	1.1 ± 1.9

Note: [a] worse group (n =19) included 13 slightly worse, 5 moderately worse and 1 very much worse

Table 4 Spearman correlation coefficients between the global rating of change and the change scores of studied PROMs

Instrument change scores	Hypothesis	spearman correlation coefficient	P-value	Supported hypothesis
VAS (0-100)	moderate positive correlation ($0.3 < r_s < 0.6$)	0.46	<0.001	yes
NRS (0-10)	moderate positive correlation ($0.3 < r_s < 0.6$)	0.46	<0.001	yes
OHIP-14 total	moderate positive correlation ($0.3 < r_s < 0.6$)	0.32	<0.001	yes
COMDQ-15 total	moderate positive correlation ($0.3 < r_s < 0.6$)	0.47	<0.001	yes
COMDQ-15-PD	moderate positive correlation ($0.3 < r_s < 0.6$)	0.4	<0.001	yes
COMDQ-15-MT	moderate positive correlation ($0.3 < r_s < 0.6$)	0.29	<0.001	no
COMDQ-15-SE	moderate positive correlation ($0.3 < r_s < 0.6$)	0.42	<0.001	yes
COMDQ-15-PS	low positive correlation ($r_s \leq 0.3$)	0.1	0.22	yes
COMDQ-26 total	moderate positive correlation ($0.3 < r_s < 0.6$)	0.49	<0.001	yes
COMDQ-26-PF	moderate positive correlation ($0.3 < r_s < 0.6$)	0.4	<0.001	yes
COMDQ-26-MT	moderate positive correlation ($0.3 < r_s < 0.6$)	0.36	<0.001	yes
COMDQ-26-SE	moderate positive correlation ($0.3 < r_s < 0.6$)	0.45	<0.001	yes
COMDQ-26-PS	low positive correlation ($r_s \leq 0.3$)	0.18	0.02	yes

Note: r_s = Spearman correlation coefficient

Table 5 Responsiveness parameter (AUC), MIC and MID estimates using different distribution-based and anchor-based methods, and recommended thresholds after triangulation process

Instruments	Distribution-based estimates		Anchor-based estimates						MIC Triangulation	MID Triangulation
	Half SD	SEM	Meaningful within-patient changes					Meaningful between-group differences		
			ROC curve analysis				mean change method	mean difference method		
			MIC	AUC	TP (%)	TN (%)				
VAS (0-100mm)	12.9	11.9	11	0.68	59	76	29	30	16	18
NRS (0-10)	1.3	1.2	2	0.69	53	84	2.7	2.7	2	2
OHIP-14 total	6.4	4.1	4	0.63	55	71	4.5	3.2	5	5
COMDQ-15 total	6	4.1	4	0.71	67	74	6.7	6.6	5	6
COMDQ-15-PD	2.3	1.6	2	0.71	69	73	3	2.5	2	2
COMDQ-15-MT	1.6	1.3	1	0.63	53	73	1.3	1.7	1	2
COMDQ-15-SE	2.6	2.4	1	0.68	73	62	2.5	2.7	2	3
COMDQ-15-PS	1.1	1.2	1	0.54	37	71	0.1	0.1	1	1
COMDQ-26 total	9.4	6.5	8	0.72	61	83	10.6	10.3	9	9
COMDQ-26-PF	3.9	2.8	3	0.71	69	72	4.7	3.8	4	4
COMDQ-26-MT	2.5	2.3	1	0.64	67	61	2.5	2.9	2	3
COMDQ-26-SE	3.6	3.2	2	0.69	65	72	3.6	3.8	3	4
COMDQ-26-PS	1.6	1.6	1	0.6	53	67	0.1	0.1	1	1

Note: Half SD = Half a standard deviation; SEM = standard error of measurement; ROC curve = receiver operating characteristic curve; MIC = minimal important change; AUC = area under the curve; TP = true positive rate; TN = true negative rate; MID = minimal important difference