Yule PL, Durham J, Playford H, Moufti MA, Steele J, Steen N, Wassell RW, Ohrbach R.  
**OHIP-TMDs: a patient-reported outcome measure for temporomandibular disorders.**  
*Community Dentistry and Oral Epidemiology* 2015, 43(5), 461-470.  

**Copyright:**  
© 2015 The Authors Community Dentistry and Oral Epidemiology Published by John Wiley & Sons Ltd  
This is an open access article under the terms of the [Creative Commons Attribution License](http://creativecommons.org/licenses/by/4.0/), which permits use, distribution and reproduction in any medium, provided the original work is properly cited.  

**DOI link to article:**  
[http://dx.doi.org/10.1111/cdoe.12171](http://dx.doi.org/10.1111/cdoe.12171)  

**Date deposited:**  
05/04/2016
OHIP-TMDs: a patient-reported outcome measure for temporomandibular disorders


This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

Abstract – Objectives: This research aims to assess the test–retest reliability, the face, content and known groups validity, and responsiveness to change, of OHIP-TMDs, a 22-item TMDs-specific version of the Oral Health Impact Profile (OHIP). Methods: Test–retest reliability – A group of patients with TMDs (n = 20) was administered OHIP-TMDs twice before initial consultation with a 2-week interval. Face and content validity – Content validity index assessments were undertaken with professionals and patients. Known groups validity – Participants (n = 76) with confirmed Axis 1 RDC/TMD diagnoses completed OHIP-TMDs prior to TMDs treatment. Their responses were compared, using inferential statistics, with those of age- and gender-matched controls. Responsiveness to change – Using the same 76 participants, a comparison was made of OHIP-TMDs with OHIP-49 (order of administration randomized) both at baseline and 3 months after starting treatment. Results: OHIP-TMDs showed good test–retest reliability ICC [2,1] 0.805 (95% CI: 0.565, 0.918); good face and content validity; significant differences (P < 0.001) between controls and participants demonstrating known groups validity. Its responsiveness to change was similar to OHIP-49. Conclusions: OHIP-TMDs is an appropriate biopsychosocial, patient-centred, outcome measure for assessing QOL in patients with TMDs. It is less than half the length of OHIP-49 and contains proportionately more items relevant to TMDs.

Temporomandibular disorders (TMDs) are a group of painful musculoskeletal conditions affecting the temporomandibular jaw joint (TMJ) and/or the muscles of mastication (1). The evidence base for the management of TMDs is problematic, and systematic reviews on TMDs have consistently, and repeatedly, called for standardized, reliable and reproducible patient-centred outcome measures for TMDs to advance and standardize trials of therapy (2-7).

There is a well-established association between TMDs symptoms, impaired general health (8) and therefore quality of life (QOL) (9-13). Generic, system-specific, or disease-/condition-specific QOL measures are available in the literature. System-specific QOL measures, for example the OHIP-49, have shown greater specificity and discriminant (known groups) validity between clinically disparate groups than generic QOL measures, for example the SF-36 (14). Condition-specific QOL measures, which are essentially further reductions/modifications of system-specific measures, such as OHIP-EDENT derived from the OHIP-49, have demonstrated less susceptibility to floor effects and a comparable responsiveness to change as the...
longer parent system-specific QOL measure (15). In addition, above a particular threshold, the reduction in the number of redundant items has been shown to improve the OHIP-49’s responsiveness to change in disease-/condition-specific situations (16).

Recent research suggests that OHIP-49 may contain a number of redundant items, which may dilute its responsiveness to change in TMDs (9, 17–19). In addition to this, generic instruments like OHIP-49 may not be sensitive to some of the factors affecting patients with TMDs (9, 20, 21). A shorter instrument is desirable for decreased statistical ‘noise’ from redundant items; administration time; complexity of scoring; cost (if used in national surveys) (15).

A shortened condition-specific form of OHIP-49, OHIP-TMDs, has recently been developed which contains twenty items from OHIP-49 and two new items derived from qualitative research with patients with TMDs, but its psychometric properties have yet to be examined (9). The aim of this study was, therefore, to validate OHIP-TMDs and assess its performance as an outcome measure for the management of TMDs against OHIP-49. The hypothesis was that in detecting changes in the impact of TMDs on quality of life, the shortened form (OHIP-TMDs) would perform as well as OHIP-49.

Materials and methods

Ethical approval was given for this study (National Research Ethics Service Committee North East 10/H0907/28), and written informed consent was obtained from all participants.

Three authors (PY, JD, RW) underwent examiner (diagnostic) calibration training with the RDC/TMD, which was periodically rechecked over the duration of the study to provide an Axis 1 diagnosis for any participant included in the study. The method used was the same as a previously published study (22).

The study had four main phases each assessing different aspects of the validity, responsiveness to change and reliability of OHIP-TMDs:

- Face and content validity of OHIP-TMDs.
- Responsiveness to change (OHIP-TMDs versus OHIP-49).
- Known groups (discriminant) validity of OHIP-TMDs.
- Test–retest reliability of OHIP-TMDs.

Sample

Participants chosen for Phase 1 were consecutive adult patients undergoing treatment for TMDs in a University Dental Hospital setting (Newcastle Dental Hospital) who were returning for review. For Phase 2, a convenience sample of consecutive new adult patients referred to the Dental Hospital with a possible diagnosis of persistent TMDs with symptoms present for at least 3 months was taken.

Putative participants for Phase 2 were identified from new patient referral letters from general dental practitioners that implicitly or explicitly suggested a diagnosis of TMDs. After telephoning to gain consent, putative participants were appointed to the research clinic and underwent a TMDs examination prior to inclusion in the study. Participants for Phase 2 were excluded if: on examination, they did not meet the RDC/TMD criteria (23) and a diagnosis of TMDs could not be made; they were found to have other significant comorbidities or orofacial pain conditions including those arising from dental disease; they could not speak or read English fluently; or they were edentulous.

Phase 3 symptomless controls were age and gender matched to the Phase 2 patients and were recruited from TMDs-free individuals who were accompanying patients in the Dental Hospital waiting rooms. A validated screening instrument (24) was used plus three other questions (10) to exclude any recent: pain, TMDs treatment or perceived treatment need for TMDs.

Phase 4 patients were recruited as a separate sample in exactly the same manner as patients recruited for Phase 2.

Instruments

OHIP-49 and OHIP-TMDs were used as self-complete questionnaires.

OHIP-49 is a 49-item problem-based questionnaire with responses recorded on a five-point, ordinal, unipolar scale: never (0), hardly ever (1), occasionally (2), fairly often (3), very often (4). The 49 items are grouped into common themes covering 7 domains: functional limitation; physical pain; psychological discomfort; physical disability; psychological disability; social disability; and handicap. OHIP-49 has been shown to have excellent psychometric properties, be responsive to change, and it and its shortened version (OHIP-14) are the only oral health-related quality of life instruments used as patient-centred outcome measures in patients with TMDs to date (10).

OHIP-TMDs is a condition-specific outcome measure for TMDs derived from OHIP-49 using a mixed-method qualitative and quantitative
approach (9). It consists of 22 items covering the same domains as OHIP-49, twenty of which came from the original OHIP-49 and two of which emerged from qualitative research with patients with TMDs (9). It records responses to the standard problem-based items of OHIP-49 on the same five-point ordinal response scale as OHIP-49 with higher scores indicating a poorer quality of life.

To calculate scores on either of the instruments, a range of methods are available (25), but no benefits have been shown of one method over another (26). The simplest method (OHIP-ADD) was used throughout this study, and it generates a summary score for the overall instrument by summing the response codes of the ordinal response scale for each item across all domains. For the purposes of the study, we have then divided the total summary score by the total number of items in the instrument to yield a score between 0 (best possible QOL) and 4 (worst possible QOL). This does not affect the psychometric properties of the scales and facilitates the direct comparison of the two instruments when necessary.

After 3 months of treatment participants also completed a single global transition judgement question (scale) to assess the degree of perceived change and as an anchor-based method of assessing the minimally important clinical difference (27). Participants were asked to score on a Likert scale whether they felt their condition had improved a lot (2), improved a little (1), stayed the same (0), worsened a little (−1) or worsened a lot (−2).

**Phases of study**

**Phase 1: face and content validity.** A focus group with participants who had RDC/TMD diagnoses and were undergoing treatment for TMDs was conducted by a trained facilitator (JD) to specifically examine the face validity of the two new questions in OHIP-TMDs and receive general feedback on the whole instrument. Discussions from the focus group were recorded digitally and anonymized. After being transcribed verbatim, they were participant to a thematic analysis (28).

Five professionals (specialist consultants in different specialties: oral medicine, oral and maxillofacial surgery, restorative dentistry) familiar with treating patients with TMDs in a secondary care Dental Hospital also independently reviewed whether all items were representative in each domain of the instrument and gave feedback on OHIP-TMDs as an instrument. Adequate representation of each domain in OHIP-TMDs was considered, and ensured, in the original mixed-methods development of OHIP-TMDs (9) consistent with the recommendations for item reduction in OHIP of Awad et al. (16).

**Phase 2: responsiveness to change.** Those who agreed to be involved after telephone contact were sent the study documentation 2 weeks in advance of their first appointment. This documentation included either OHIP-TMDs or OHIP-49 to be completed and sent back prior to attending the clinic. A permuted block randomization procedure was used to determine which questionnaire was administered first. When the patient attended for their first clinic appointment, they then completed the other questionnaire in the waiting room prior to seeing the clinician. Patients therefore had 2 weeks between completing each of the questionnaires (OHIP-49 and OHIP-TMDs) over which little change in their condition was likely to occur, but allowed sufficient time to be unlikely to recall previous responses and bias their response to the questionnaire administered second. Clinicians were blinded both to the order of the questionnaires and their contents throughout the duration of the study. The association between OHIP-TMDs and OHIP-49 scores, including the impact of questionnaire order, was investigated using regression analysis.

Following examination and diagnosis, the patient was given an explanation of the condition, advice on self-help and initial reversible therapy selected as appropriate by the clinician. A review was booked at 3 months to allow sufficient time for a clinically important change to occur (29). Two weeks prior to the review appointment, the OHIP-49 or OHIP-TMDs were similarly re-administered with the first questionnaire sent to the patient in the post and the second questionnaire administered in the waiting room prior to the review appointment. At the review appointment, the patients were also asked to rate the global change in their condition using an anchor-based method, in line with IMMPACT recommendations (27, 30), on a global transition judgement scale (GTJS). Participants who failed to attend the review session or failed to complete their questionnaires were followed up by phone to encourage them to complete missing questionnaires and return them by post. Those not responding after two telephone calls were dropped from the study.

There is little agreement in the literature on sample sizes for determining responsiveness to change.
(31). Previous research assessing the responsiveness of other instruments (32) has used relatively small sample sizes of 40–50 patients observing moderate effect sizes ranging from 0.34 to 0.67 for the summary scores of different instruments whose intraclass correlation coefficients were in the acceptable range of 0.59–0.93 (31, 32). We a priori set our sample size for Phase 2 by examining the effect sizes with respect to OHIP-49 from previously unpublished data assessing responsiveness to change during treatment (29). The effect size and its confidence intervals were 0.52 (95% CI: 0.27–0.76). As we would expect OHIP-49 to be less responsive to change than OHIP-TMDs, and given the intended purpose of the instrument in being able to detect the smallest effect size required to show a change in a patient’s quality of life, we chose a moderate effect size of 0.35 for Cohen’s d (33) from the lower end of the confidence interval for OHIP-49. This therefore meant that for a power of 80% (α = 0.05) we required 67 patients to complete both questionnaires pre- and post-treatment. To allow for an expected dropout of 33%, as experienced by similar studies (32), we aimed to recruit a minimum of 100 participants.

Phase 3: known groups validity. People accompanying patients in the Dental Hospital waiting rooms were approached and invited to take part in Phase 3 of the study. These controls were age and gender matched to the participants in Phase 2. Screening questions (10, 24) were used prior to the participants completing OHIP-TMDs to ensure they were TMD free, did not have any other painful orofacial condition and had never had any symptoms of, or treatment for, TMDs.

Phase 4: test–retest reliability. The participants for Phase 4 were recruited in the same way as Phase 2. Each participant completed OHIP-TMDs twice with a 2-week period between administrations. It was determined that between 8 and 23 participants would be required to complete Phase 4 to obtain at least 80% power for detecting an intraclass correlation coefficient of between 0.6 and 0.8 ['substantial agreement' (34, 35)]. Again to allow for an expected 33% dropout, we aimed to recruit a minimum of 30 participants.

Data analysis
There were no missing data from the control participants or the participants in Phase 4. Missing data were found in <10% of items from each participant in Phase 2. No participant was therefore excluded. Imputation of any missing data followed accepted and previously published methods for OHIP (10, 36). The most common missing items were logically missing items, given dentate status, relating to problems with denture fitting and denture comfort. Once these were accepted as logical and the proportion of missing items per questionnaire (OHIP-TMDs baseline 0.12% missing items, follow-up 0%; OHIP-49 baseline 0.08%, follow-up 0%) were compared, there was no significant difference (P > 0.05, test of proportions) in the proportion of missing items between the two questionnaires.

Data were entered into Excel (Excel v10, Microsoft Corporation, Redmond, WA, USA) spreadsheets and then transferred to the statistical package STATA (STATA/IC 12 Statistical Software, StataCorp LP, College Station, Texas, USA) for analysis.

Phase 1 – face and content validity. The face validity of OHIP-TMDs was specifically examined in a focus group of TMD participants uninvolved with any other phase of the study, and data from this group were participant to thematic analysis. The content validity of OHIP-TMDs was assessed using the content validity index with the focus group participants and the health care professionals independently. The content validity index required each item of OHIP-TMDs to be rated by each participant and by each professional as to its relevance to TMDs using an ordinal rating scale where 1 = not relevant at all, 2 = not really relevant, 3 = quite relevant, 4 = highly relevant. The content validity index values were then collapsed to either the individual agreeing (3,4) or not agreeing (1,2) the relevance of the item. Those agreeing were scored 1, and those not agreeing its relevance were scored 0. Each item therefore had a content validity score, and these scores could be summed, for each domain, and also for the entire questionnaire with a higher mean score giving an indication of greater, perceived, content validity (28).

Phase 2 – responsiveness to change. Paired, two-tailed, t-tests were completed on the data to examine for significant differences between administrations of the questionnaires. The magnitude of change of the instruments was measured by calculating the effect size (ES, Cohen’s d) (15, 37) for each instrument. Confidence intervals for the ES were calculated using bootstrapping employing
1000 repetitions and the bias-corrected and accelerated technique (38).

The minimally important clinical difference was also calculated using methods described by Allen et al. (39) using the global transition judgement as the external criterion of change (anchor). Linear regression was also completed against the global transition judgement scale.

Cronbach’s alpha was used as a co-efficient of reliability to measure the internal consistency of the instruments.

**Phase 3 – known groups validity.** Independent measure t-tests of the mean item scores for the control participants against the mean item baseline scores of Phase 2 patients completing OHIP-TMDs were used to determine the known groups validity of OHIP-TMDs.

**Phase 4 – test–retest reliability.** Test–retest reliability of OHIP-TMDs was determined by calculating the intraclass correlation coefficient ICC (2,1) using a two-way random effects model with absolute agreement according to the Shrout and Fleiss convention (40).

**Results**

**Face and content validity (Phase 1)**

Eight patients were recruited to the focus group, with five attending on the day. The remaining three either failed to attend \( n = 2 \) or cancelled due to ill health \( n = 1 \). The patients ranged from 18 to 69 years and had a wide range of TMDs diagnoses (Table S1: e-appendix). The results of the thematic analysis of the patient focus group suggested a number of small modifications should be made to OHIP-TMDs before it was used for the main parts of the study (Phases 2,3,4, final version available in e-appendix): changed all items with the words ‘jaws, teeth, mouth or dentures’ to ‘jaws, teeth or mouth’; changed item 3 from ‘painful aching in your mouth’ to ‘painful aching in your mouth, face or ear’; changed item 7 from ‘have you felt speech was painful’ to ‘have you felt talking was painful’.

The patients thought there was ambiguity in using the word ‘dentures’ as denture users may have non-TMDs complaints and agreed the changes to the items as detailed above. There were no issues raised with the two items in OHIP-TMDs that were not from OHIP-49 and therefore had not had their face validity assessed previously.

In addition to the patients involved in the focus group, five professionals were also independently involved in assessing the content validity and cross-checking the patient’s suggested changes for face validity as above. There were no discordant responses from the professionals regarding face validity changes suggested by the focus group.

The content validity index values per domain for this phase are shown in Fig. S1: e-appendix. The values per domain for the professionals were generally closer to 1.0 than the patient’s values, suggesting the professionals felt more of the items were relevant to TMDs. The overall questionnaire mean content validity index score for the patients was 0.64 and for the professionals 0.82. When the patient and professionals results were combined, the mean content validity index score for the overall questionnaire was 0.73.

The two items where patients and professionals showed complete agreement on their relevance to TMDs were as follows:

- ‘Have you had difficulties opening or closing your mouth?’
- ‘Have you had a sore jaw?’

**Responsiveness to change (Phase 2)**

A total of 139 patients were invited to take part in this Phase of the study, with 76 completing it (Recruitment flow diagram in Fig. S2: e-appendix).

There were 17 patients who failed to complete the study (classed as noncompleters). The mean age of these patients was 38.5 years (SD = 14.9). There was no significant difference between those patients that completed the study and those that did not in terms of which questionnaire they received first.

The mean age of the 76 patients who completed the study was 44.9 years (SD = 15.3), with 86% (65) of them being women. They had a wide range of RDC/TMD diagnoses (Table S1: e-appendix). Just less than half of the sample (43%) had received previous treatment for TMDs with most of this being provided in primary care as a soft splint (79%). A variety of conservative TMDs management was provided within the study for the whole sample (Table S2: e-appendix).

In general, mean OHIP-TMDs scores were larger than mean OHIP-49 scores (Fig. S3a: e-appendix). The estimated mean difference between OHIP-TMDs and OHIP-49 summary scores based on a bootstrap procedure was 0.57 (95% CI: 0.47, 0.66). Inspection of the plot (Fig. S3a: e-appendix) suggests that any effect of questionnaire order is likely to be modest. This was further investigated
Table 1. Phase 2 mean change scores and effect sizes for OHIP-49 and OHIP-TMDs

<table>
<thead>
<tr>
<th>Instrument</th>
<th>Mean raw change score (SD)</th>
<th>Mean change per item (SD)</th>
<th>Effect size (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>OHIP-49</td>
<td>11.7 (± 19.4)</td>
<td>0.2 (± 0.4)</td>
<td>0.4 (0.3, 0.6)</td>
</tr>
<tr>
<td>OHIP-TMDs</td>
<td>6.9 (± 15.9)</td>
<td>0.3 (± 0.7)</td>
<td>0.4 (0.2, 0.6)</td>
</tr>
</tbody>
</table>

by modelling the relationship between mean OHIP-TMDs and OHIP-49 summary scores. A linear relationship between the two instruments over the full range of possible mean summary scores (0–4) is clearly not plausible. A second order polynomial constrained to pass through (0,0) and (4,4) yielded a reasonably fitting model (R² = 0.65). Based on this model, the estimated mean OHIP-TMDs summary score is plotted as a function of the mean OHIP-49 summary score in Fig. S3b: e-appendix. Adding an effect of questionnaire order to the regression model did not yield a significant improvement in fit; the estimated impact of questionnaire order on the difference between mean OHIP-TMDs and OHIP-49 summary scores was 0.01 [95% CI: −0.02, 0.05].

Table 1 shows the mean change scores and the effect sizes for OHIP-TMDs and OHIP-49 following treatment. For both instruments, the change in score was significant (P < 0.001). The minimally important clinical difference (MICD) is the smallest difference (or improvement) in their condition which patients perceive as being beneficial (41). Here the MICD was calculated as the mean change in questionnaire score for the GTJS category ‘improved a little’ (39) as shown in Table 2. Table 3 shows the adjusted coefficient of determination for change in OHIP-49 and OHIP-TMDs within a linear regression model against the GTJS examining whether the patients’ condition had ‘worsened a little’, ‘stayed the same’, ‘improved a little’, improved a lot’ following treatment. The gradient of the line, given by the regression coefficient, corresponds to the estimated change in questionnaire scores between each category of the GTJS.

Cronbach’s alpha was 0.95 for the baseline administration of OHIP-TMDs and 0.96 for the follow-up administration of OHIP-TMDs, which was identical to the reliability coefficients for baseline, and follow-up, administration of OHIP-49.

Known groups validity (Phase 3) There were 76 controls recruited that were age and gender matched to the patients completing Phase 2. The mean age of the controls was 46.7 years (SD = 17.8), with 86% (65) of them being women. The mean OHIP-TMDs summary score for the controls was 7.38 (SD = 10.1) compared to 33.4 (SD = 17.07) for the patients at baseline. Each individual item in OHIP-TMDs was significantly different in score between patients and controls (P < 0.00).

Test–retest reliability (Phase 4) Twenty patients completed this test–retest Phase with a mean age of 41.9 years (SD = 20.1). A total of 85% (17) of these patients were women. Again patients had a wide range of TMDs diagnoses (Table S1: e-appendix). The majority (85%) had had previous treatment for TMDs, with all but one patient having been seen in primary care for this. Most of these patients (15; 88%) had been prescribed a soft splint in the past.

The difference in OHIP-TMDs mean summary scores between baseline and pretreatment follow-up in this group of patients was 32.4 (SD = 21.9) and 27.9 (SD = 17.5), respectively. The reliability of the observations for the whole instrument – intra class correlation coefficient (2,1) – was 0.805 (95% CI: 0.565, 0.918).

For purposes of comparability, the treatment that Phase 2 and Phase 4 patients received after being examined at Newcastle Dental Hospital is described in Table S2: e-appendix.

Discussion

It is important that TMDs are studied from a biopsychosocial perspective. That is, the biological or

Table 2. Phase 2 mean change score of questionnaires by global transition judgement scale category

<table>
<thead>
<tr>
<th>Instrument</th>
<th>Global transition judgement score</th>
<th>−1 ‘worsened a little’</th>
<th>0 ‘stayed the same’</th>
<th>+1 ‘improved a little’</th>
<th>+2 ‘improved a lot’</th>
</tr>
</thead>
<tbody>
<tr>
<td>OHIP-49</td>
<td>Mean change (SD)</td>
<td>−7 (± 16.2)</td>
<td>−1.2 (± 14.2)</td>
<td>12.5 (± 16.9)</td>
<td>24.5 (± 17.9)</td>
</tr>
<tr>
<td></td>
<td>Mean change per item (SD)</td>
<td>−0.1 (± 0.3)</td>
<td>0 (±0.3)</td>
<td>0.3 (± 0.3)</td>
<td>0.5 (± 0.4)</td>
</tr>
<tr>
<td>OHIP-TMDs</td>
<td>Mean change (SD)</td>
<td>−9.9 (± 5.1)</td>
<td>−1.9 (± 10.8)</td>
<td>6.9 (± 13.1)</td>
<td>17.6 (± 15.7)</td>
</tr>
<tr>
<td></td>
<td>Mean change per item (SD)</td>
<td>−0.5 (± 0.3)</td>
<td>−0.8 (± 0.5)</td>
<td>0.3 (± 0.6)</td>
<td>0.8 (± 0.7)</td>
</tr>
</tbody>
</table>

MICDs shown in bold.
The biomedical concept of how the physical disorder of TMDs is linked to the psychological and social concepts of how TMDs affect the individual (42, 43).

The biopsychosocial perspective matters because it is now widely accepted that TMDs are persistent, recurrent painful conditions, which can be self-limiting and are associated with ‘appreciably distressful, but typically non-specific clinical symptoms’ (42, 43). As such, treatment outcome for TMDs needs to focus not just on pain reduction, but also on improving patients’ quality of life.

The negative impact on individuals’ quality of life by TMDs can be captured by OHIP-49 (13, 18, 19, 44). OHIP-49 is, however, a long questionnaire with a number of redundant items that may reduce its responsiveness for TMDs studies (9, 10). Systematic reviews have repeatedly emphasized the need for a single, condition-specific instrument assessing the quality of life of patients with TMDs. Such an instrument would help reduce heterogeneity between TMDs trials, allowing more effective meta-analyses and thereby facilitate changes in patient management based on best available evidence.

The new instrument, OHIP-TMDs, validated in this study generally showed good face and content validity at both the individual item level and for the whole instrument. Previous research has indicated the standard criterion for acceptability of the mean content validity index score for a whole instrument is 0.8 (28, 45, 46) or 0.9 (47), the overall result in our study being only slightly less than this. The mean content validity index score for the professionals was above the standard (0.82), with that of the patients being below the standard (0.64). There was little agreement generally between the professionals and the patients as to which items were quite or highly relevant to TMDs. This finding is unsurprising as professionals are likely to have a better overall understanding of TMD symptoms, compared to patients who are likely only to be able to relate to a specific item if it maps to a symptom, they have individually experienced. Furthermore, patients with the same physical TMD diagnosis may have differing levels of psychosocial impact (48), and therefore, despite the same physical diagnosis, the different psychosocial diagnosis or impacts may make it difficult for seemingly homogenous patients to agree a consensus on how TMDs affect their QOL.

Our sample of patients with TMDs chosen to evaluate responsiveness to change had a broad spread of TMDs diagnoses (Table S1: e-appendix) and received a variety of treatments (Table S2: e-appendix). Regarding responsiveness to change, the higher magnitude of change with OHIP-49 is to be expected bearing in mind this instrument has more than twice as many items as OHIP-TMDs. However, OHIP-TMDs performed as well as OHIP-49 in terms of its effect size and associated confidence intervals (Table 1). Treatment outcome (Table 2) was also assessed globally using a GTJS in terms of the patients’ perceptions of whether their condition had improved, stayed the same or worsened since starting this latest episode of treatment. Of the patients reviewed after 3 months, 53 (70%) reported an improvement with an almost equal division between those who improved slightly and those who had much improved. Only seven patients (9%) reported a worsening of their condition whilst 16 (21%) reported no improvement. Whilst a GTJS has the disadvantage of recall bias due to its retrospective nature (49), it does, however, generate a single score for comparison against other instruments. In our study, it is clear that the mean score for the OHIP-TMDs and OHIP-49 within each of the five GTJS categories had the expected relationship. This was confirmed with the significant regression line gradients for both instruments (Table 3).

The concept of the minimal important clinical difference (MICD) is one which has been used frequently in medicine in combination with generic and disease-specific instruments (50), particularly for assessing chronic pain. It has also been defined for oral applications, for example for OHIP-20 and OHIP-14 (32, 39). It is the smallest difference in scores which patients perceive as either being beneficial or as representing deterioration in their condition (39). The MICD was determined for this heterogenous group of patients with TMDs using a recognized method described by Allen et al. whereby the mean OHIP change score is compared to GTJS scores for those patients who scored themselves as ‘slightly improved’ (39). Whilst the MICD
may indeed vary with different TMDs diagnoses, within either Axis I or II, the MICD score derived from this study provides an initial estimate for either instrument which could be applied in clinical trials to allow for analysis of patients by the proportion of patients who exceed the MICD for any particular treatment (51, 52). This approach allows the ‘number needed to treat’ to be calculated (53) for each treatment group. The ‘Clinically Important Difference’ (CID) calculated as the mean of the ‘much improved group’ would give a more rigorous assessment of treatment outcome. However, rigour should be tempered by realism as TMDs can be chronic conditions where patient management emphasizes improvement rather than cure.

As can be seen from the gradient of the lines of the regression analyses, these are close to the MICD value calculated using Allen’s method, as described above, for the two instruments, providing a further estimate of the MICD and the CID.

Other methods, including receiver operator curves (ROC) (52), have been described to calculate MICD. The ROC method works well with large patient groups and could be used in future studies to corroborate or refine our findings, remembering that the primary aim of this study was to compare OHIP-49 with OHIP-TMDs.

OHIP-TMDs showed excellent known groups validity and good test–retest reliability. As regards the latter, the ICC showed moderate to high levels of reliability, with a large confidence interval around the ICC. This confidence interval is probably expected as the signs and symptoms experienced by patients with TMDs often fluctuate (54), especially prediagnosis and pretreatment.

The main limitations of this study relate predominantly to the convenience sample, which reflected the normal bias towards women in both clinical and research settings. It is known that responses to OHIP can differ based on culture and country (36, 55) although in a primarily English speaking culture, we would hope that OHIP-TMDs would be largely transferable. We also did not formally conduct an Axis 2 RDC/TMD examination and therefore cannot directly examine the cohort’s psychosocial characteristics, but we feel given the range of responses in the psychosocial domains of OHIP that our cohort has sampled the broad range of presentations.

There was a relatively large difference in the number of patients invited to take part in the study, compared to those who completed the study. This is likely due to the fact the design of the study required a high level of compliance from patients. Locker et al., (32) similarly had a large sample size as they expected the numbers of drop-outs to be high.

This study has highlighted that OHIP-TMDs and OHIP-49 perform similarly when used as a TMDs outcome measure. The choice of instrument relates to its intended purpose. This said, OHIP-TMDs has less than half the number of items than OHIP-49 and therefore has the advantage of reducing response burden and potential acquiescence and attrition bias with cohorts of patients with TMDs. OHIP-TMDs would also reduce the time involved in completing instruments if used in a busy clinical setting. Hence, OHIP-TMDs provides a more streamlined outcome measure than OHIP-49 for TMDs treatment trials or evaluation studies. This more pragmatic approach of using shorter instruments has been advocated by other authors (15, 56).

OHIP-14 is a valid outcome measure to assess oral health-related quality of life, but it is not a condition-specific measure for TMDs and is therefore not ideal as it is not balanced enough or domain representative for TMDs (57, 58). OHIP-TMDs includes all seven domains covered by OHIP-49 (9) and like OHIP 49 it has been translated into languages other than English, with the recent validation of a Chinese version OHIP-TMDs-C (59). It is important to note, however, that the recent dimensions of oral health-related quality of life project have suggested that only the summary score be used for all versions of OHIP because of methodological weaknesses in the domain scoring system (60).

As a final word, the new QOL outcome measure would of course supplement and in no way replace accepted pain measurements in clinical trials.

**Conclusion**

The new patient-centred outcome measure, OHIP-TMDs, showed good content, face and known groups validity. It was also responsive to change and had good test–retest repeatability. OHIP-TMDs and OHIP-49 were very similar in their effect size and magnitude of change. Either OHIP-49 or OHIP-TMDs are appropriate instruments to measure QOL as a treatment outcome, but OHIP-TMDs has greater utility being shorter and containing items specifically relevant to TMDs.
Acknowledgements

This research was funded by the Royal College of Physicians and Surgeons of Glasgow TC White Young Researcher Grant. The authors have no conflict of interests to report in relation to this research. We are grateful to Dr Nigel Adams, former Trust Grade Dentist at Newcastle Dental Hospital, who made a substantial contribution in setting up the study and to Dental Nurses Angela Fenwick and Elizabeth Smith for their clinical help. Finally, we would like to express our deep gratitude to the patients and health professionals who were involved in this study.

References


**Supporting Information**

Additional Supporting Information may be found in the online version of this article:

**Fig. S1.** Mean collapsed content validity index scores per domain of OHIP-TMDs.

**Fig. S2.** Flow diagram of patients recruited and enrolled to Phase 2 of the study.

**Fig. S3.** (a) Mean summary scores dependent on questionnaire order. (b) Estimated (predicted) mean OHIP-TMDs score plotted as a function of the mean OHIP-49 score.

**Table S1.** Summary of the sample TMD Diagnoses of patients in Phases 1, 2 and 4, as per the RDC/TMD Groupings, expressed as a percentage of the total number of patients in that Phase. Patients could have more than one diagnosis.

**Table S2.** Treatment prescribed for patients in Phase 2 (n = 76) and Phase 4 (n = 20) per RDC Grouping.

**Appendix S1.** Full final version of OHIP-TMDs.