Psychological screening of temporomandibular disorder patients

van der Meulen, M.J.

Publication date
2011

Chapter 6

Developing abbreviated OHIP versions for use with TMD patients

M. J. van der Meulen, M. T. John, M. Naeije, F. Lobbezoo

Journal of Oral Rehabilitation (in press)
2011 Aug 9 [Epub ahead of print]

Abstract

The aim of this study was to evaluate the psychometric characteristics of three versions of the Dutch Oral Health Impact Profile (OHIP-NL), for clinical use with temporomandibular disorder (TMD) patients. To that end, two abbreviated OHIP versions (OHIP-NL14 and OHIP-NL5) were developed by respectively selecting 14 and 5 items from the officially translated and culturally adapted original 49-item OHIP-NL questionnaire. 245 consecutive patients, referred by their dentist to the TMD clinic of the Academic Centre for Dentistry Amsterdam (77 % female; mean age ± SD = 41.0 ± 14.9 years), completed the Research Diagnostic Criteria for TMD (RDC/TMD) axis II questionnaire and the OHIP-NL. Reliability and validity of all three OHIP versions were compared, and their associations with four psychological axis II variables, indicating the TMD patients' level of impairment, were examined. According to guidelines for clinical application, internal consistency scores were sufficient for OHIP-NL and OHIP-NL14, but insufficient for OHIP-NL5. Test-retest reliability (n=64) was excellent for OHIP-NL, and OHIP-NL14 and fair to good for OHIP-NL5. For all three versions, there was evidence for score validity: associations between OHIP summary scores on the one hand and validation variables and other RDC/TMD axis II variables on the other hand met the expectations and were statistically significant ($P < 0.001$). In conclusion, the OHIP-NL and OHIP-NL14 both performed comparatively well, and better than the OHIP-NL5. When the length of the questionnaire (i.e., the time needed for its completion) is an issue, the OHIP-14 would therefore be the preferred version.

Introduction

The Research Diagnostic Criteria for Temporomandibular Disorders (RDC/TMD) (1) is the most widely used instrument for the diagnosis of TMD. In the RDC/TMD, a dual axis classification is recommended. On axis I, physical complaints are classified, while axis II relates to the psychological and behavioral aspects of TMD. One of the axis II instruments is the Graded Chronic Pain (GCP) scale. The two factors that determine the GCP classification are pain intensity and pain-related disability, indicating the interference of TMD pain with the patient's daily life (2).

High levels of pain-related disability have been suggested to be related to the development of chronicity of pain complaints and to poor treatment outcome (3). Based on the results of several studies in which the impact of different oral problems on quality of life was examined (4-6), the concept of oral health-related quality of life (OHRQoL) was found to be associated with pain-related disability, both possibly measuring two related constructs. During a recent workshop of the International RDC/TMD Consortium Network, which aimed at modifying the RDC/TMD into Diagnostic Criteria (DC/TMD) for routine clinical implementation and at introducing new measures for the updated RDC/TMD versions, it was decided to include a measure of OHRQoL (7). The instrument suggested for that purpose was the Oral Health Impact Profile (OHIP); a 49-item questionnaire, which was developed by Slade and Spencer in 1994 (8). The OHIP as a global instrument for measuring perceived oral health would provide an opportunity to compare the patient-perceived burden from oral conditions not only across TMD settings, but also across other dental settings. The instrument has been translated in different languages, and has good to excellent psychometric properties (e.g., 9-14).

For practical purposes, the use of a more succinct instrument was considered preferable. To that end, several abbreviated versions were derived from the original OHIP-49 (15-23). The OHIP selection for the present study was based on length and comparability. Two abbreviated versions were selected, namely the first developed and most widely used OHIP with 14 items (15) and the shortest OHIP with 5 items (16), to be examined in the Dutch cultural environment. Together with the long OHIP, these two short versions make OHRQoL assessable in all settings, even in those where patient burden is a concern. This is particularly important for the field of TMD, where a number of self-report instruments are already recommended to assess the patient comprehensively (7). While the OHIP could be used in addition to, or instead of other instruments, patient burden has to be minimized. On the other hand, comparability of results should be maximized, following the primary goal of the International RDC/TMD Consortium Network (1) to facilitate international comparisons between researchers. Because psychometric properties for abbreviated Dutch-language OHIP versions are lacking, it was the aim of this study to compare the psychometric properties of two abbreviated versions of the Dutch OHIP with the original 49-item OHIP-NL questionnaire (14) in a sample of TMD patients.

Materials and methods

Subjects, setting, and study design

A sample of consecutive patients, who were referred by their dentist to the TMD clinic of the Academic Centre for Dentistry Amsterdam because of TMD as their primary complaint, participated in this study (n=245; 77 % female; mean age ± SD = 41.0 ± 14.9 years). Before their first appointment with the dentist, all patients completed the axis II questionnaire of the Research Diagnostic Criteria for TMD (RDC/TMD) (1) as well as the 49-item OHIP-NL (14).

Development of abbreviated OHIP versions

The abbreviated OHIP versions were extracted from the 49-item OHIP-NL. The OHIP-49 consists of 49 questions that were conceptually categorized according to a hierarchical model, in which different levels of disease-related disruptive impairment were distinguished. Seven domains were identified. The domain 'Functional limitation' concerns the loss of function of parts of the body, like difficulty with chewing. The domains 'Physical discomfort' and 'Psychological discomfort' deal with experiences of pain and discomfort, such as toothache and feeling miserable. The domains 'Physical disability', 'Psychological disability', and 'Social disability' refer to limitations in performing daily life activities, like avoiding certain foods, lack of concentration, and feeling irritable with others, respectively. Finally, the domain 'Handicap' concerns a sense of disadvantage in functioning, like suffering financial loss due to dental problems.

Several abbreviated versions were derived from the original OHIP-49 (15-23). Except for language differences, variations in the short OHIP study designs included applying different statistical techniques to select items, developing OHIP versions for special patient groups, varying the number of items used, involving patients in the selection of items, letting subjects give weights to items they thought were more important than others, and using different scoring techniques, like summary, weighted, or standardized scores. The abbreviated OHIP versions selected for the present study were based on two already existing short OHIP versions, namely the English-language OHIP-14 (15) and the German-language OHIP-5 (16). The items of the OHIP-14 were selected based on two main conditions: 1. each of the original seven domains had to be represented with at least one question; and 2. those questions were selected that had the highest explained variance of the OHIP-49 total score (15). For the development of the German OHIP-5, which was set up for situations where lack of time would only allow for a very short questionnaire, the selection of items was based on the criterion that 90% of the variance of the OHIP-49 summary score should be explained by the selected items (16). For the Dutch versions of the OHIP-14 and -5, the corresponding items from the OHIP-NL (14) were selected. The OHIP-NL's reliability and validity were shown to be good to excellent with a group of prosthodontic and implantology patients (14). Answers to the oral health impact questions were scored on a 5-point ordinal scale, ranging from 'Never' [0], 'Hardly ever' [1], 'Occasionally' [2], and 'Fairly often' [3], to 'Very often' [4]. The total scores were the sums of the item responses and could thus range from 0 to196, from

0 to 56, and from 0 to 20 for the OHIP-NL, the OHIP-NL14, and the OHIP-NL5, respectively.

The abbreviated instruments' psychometric properties were evaluated by examining their reliability and validity and comparing these with the OHIP-NL.

Reliability

For the OHIP-NL, the OHIP-NL14, and the OHIP-NL5, internal consistency was assessed, using Cronbach's alpha (24) and average inter-item correlations. Following Bland and Altman (25), Cronbach's alphas of 0.70-0.80 were considered satisfactory. These authors state that for clinical application, much higher alpha values are needed. The minimum is 0.90, while alpha = 0.95, is desirable. Mean inter item correlations were set at 0.40-0.50 (26)

Test-retest reliability was assessed in a convenience sample of 64 consecutive TMD patients, using a time interval of two weeks between administration of the two OHIP-NL questionnaires. In those two weeks, patients did not receive any treatment. Intraclass correlation coefficients (ICCs) were calculated, based on one-way repeated measures ANOVA (27), with subjects as the factor of interest. In addition, the method of Bland and Altman (28) was used to quantify test-retest differences, and to detect systematic differences between scores on the two questionnaires. ICC values < 0.40 were considered to show poor reliability; values ≥ 0.40 and ≤ 0.75 were considered fair to good; and values > 0.75 were considered excellent (29).

Validity

The validity of the OHIP-NL14 and OHIP-NL5 was established and compared with the OHIP-NL in three ways:

1. Since the construct validity of the OHIP-NL was already established (14), we used the long instrument as the criterion measure and compared the short OHIPs' scores with the original OHIP-NL. Pearson correlation coefficients were calculated for the correlations between the instruments' summary scores. Correlations taking the part-whole correlation between the two measures into account were also calculated (30).
2. For convergent validity, the relationships between scores on the OHIPs and measures that are assumed to be derived from related constructs were examined. The following measures were used:
    - Pain-related disability score, determined by combining three questions: "In the past six months, how much has your facial pain interfered with 1. your daily activities; 2. your ability to take part in recreational, social and family activities; and 3. your ability to work, including housework, rated on a 0 to 10 scale where 0 is 'no interference' and 10 is 'unable to carry on any activities'?". The pain-related disability score was computed from those three questions. Scores ranged from 0-100 (average score times 10).
    - Self-reported oral health status, which was established with a single question: "Would you say your oral health status in general is…?". The answers ranged from 'Excellent' [1], 'Very good' [2], 'Good' [3], and 'Fair' [4], to 'Poor' [5].

3. For group validity, OHIP scores were compared between patients with or without complaints that had been shown to be related to OHRQoL, and between patients with or without oral conditions that were expected not to be related to OHRQoL.

Expected to be related to OHRQoL were:

- RDC/TMD axis II Characteristic Pain Intensity, namely the mean score of three questions: "How would you rate 1. your facial pain at the present time; 2. your worst pain; and 3. your average pain, on a 0-10 scale, where 0 is 'no pain' and 10 is 'pain as bad as could be'?". The score could range from 0-100 (mean score times 10).
- Self-report of oral conditions: "Do you sometimes have a burning sensation in your mouth?". This question could be answered 'No' [0] or 'Yes' [1].

Expected not to be related to OHRQoL were:

- Joint clicking: "Does your jaw click or pop when you open or close your mouth or when chewing?", to which the answer could be 'No' [0] or 'Yes' [1].
- Biting activities. The mean score on three questions was used: "How often during the past time have you engaged in the following activities?": 1. nail biting; 2. biting on pen/pencil; and 3. chewing gum. Answers could be 'Never' [0], 'Sometimes' [1], 'Regularly' [2], 'Often' [3], and 'Always' [4].

It was hypothesized that subjects with lower pain-related disability scores, better self-reported oral health, less TMD pain, or no burning mouth sensations would have lower OHIP-NL, OHIP-NL14, and OHIP-NL5 scores. It was also expected that the presence of joint clicking and of biting activities would have no substantial influence on OHRQoL. It is an important principle of group validity to demonstrate that the new instrument is not universally associated with oral health conditions. It should not be associated with conditions where an impact on OHRQoL is expected to be absent a priori.

When measures were continuous, e.g., pain-related disability scores, tertiles were formed to show whether OHIP summary score means follow the predicted direction. Spearman rho correlations were calculated to examine the associations between the OHIP summary scores on the one hand and tertiles of pain-related disability scores, characteristic pain intensity, and biting activities on the other hand. T-tests were performed to examine the associations between the OHIP summary scores on the one hand and burning mouth sensations and joint clicking on the other. Correlations between 0.10 and 0.29 were considered as low, between 0.30 and 0.49 as moderate, and equal to or higher than 0.50 as high (31).

Levels of impairment for the RDC/TMD axis II measures: Associations between OHIP scores and RDC/TMD axis II measures

Following John's model (6), associations were determined between the summary scores of all three versions of the OHIP and the main psychological RDC/TMD axis II measures:

- Graded Chronic Pain (GCP), which classifies patients according to their pain intensity and pain-related disability, and ranges from 0 to 4. In this classification, two measures of pain-related disability, namely the disability score and disability days are used. For the latter, one question is used, concerning the number of days over the past six

months that the patient is kept from usual activities (work, school, or housework) because of facial pain. Both disability score and disability days may then each result in disability points (ranging from 0 to 6), indicating the severity of the pain-related disability. The GCP classification distinguishes between five subgroups: a pain-free grade (0); a low disability/low pain intensity grade (CPI <50) (I); a low disability/high pain intensity grade (CPI ≥50) (II); a high disability, moderately limiting grade (3 or 4 disability points), regardless of pain intensity (III); and a high disability, severely limiting grade (5 or 6 disability points), regardless of pain intensity (IV).

- Jaw disability, for which the Jaw Disability Checklist (1) was used. The question was formulated as follows: "What activities does your present jaw problem prevent or limit you from doing?". It includes 12 items, like chewing, yawning, and talking. The answer to each item can be 'No' [0] or 'Yes' [1]. Total scores can range from 0 to 12 (1).
- Nonspecific physical symptoms and depression, measured with the Somatisation and Depression scales of the Dutch version of the Symptom Checklist 90 (SCL90) (32). Questions relate to the degree in which the subject has been bothered by symptoms during the past week, including today. The questionnaire includes 12 Somatisation items and 16 Depression items. Responses may range from 'Not at all' [1], 'A little' [2], 'Somewhat' [3], and 'Rather much' [4], to 'Always' [5]. Total scores may range from 12 to 60 for Somatisation, and from 16 to 80 for Depression.

It was hypothesized that subjects in lower numbered GCP subgroups, with lower jaw disability, and with lower SCL90 scores would have lower OHIP-NL, OHIP-NL14, and OHIP-NL5 scores. Spearman rho correlations were calculated to examine the associations between the OHIP summary scores on the one hand and Graded Chronic Pain, tertiles of jaw disability, and SCL90 scores, on the other hand.

Missing data

Two out of the 245 subjects, who missed more than five questions on the OHIP-NL, or more than two questions from within one of the seven domains, were discarded from the statistical analyses. Missing answers that did not exceed these criteria, were imputed using imputation within the relevant domain, i.e., the domain's mean was calculated and entered for missing values.

All analyses were performed using the statistical software package STATA, Release 9 (StataCorp. 2005, Stata Statistical Software, College Station, TX, USA), with the probability of a type I error set at the 0.05 level.

Results

Reliability

Cronbach's alpha's and average inter-item correlations of the OHIP summary scores are shown in Table 1. The internal consistency of OHIP-NL and OHIP-NL14 exceeded the

minimal standard value of "satisfactory", however the OHIP-NL5 just missed that standard. For clinical application, the internal consistency of the OHIP-NL was excellent and that of the OHIP-NL14 at the required level, but for the OHIP-NL5, it was below the minimum required level. For test-retest reliability (n=64), intraclass correlations, mean differences, and limits of agreement are shown in Table 1. The test-retest reliability was excellent for the OHIP-NL and the OHIP-NL14, and fair to good for the OHIP-NL5.

Table 1. Internal consistency (Cronbach's alphas and average inter-item correlation) and test-retest reliability (Intraclass correlation coefficients (ICCs; mean difference; and limits of agreement) of the OHIP-NL, OHIP-NL14, and OHIP-NL5

OHIP version	Internal consistency		Test-retest reliability		
	Cronbach's α	Average inter-item correlation	ICC	Mean difference	Limits of agreement
OHIP-NL	0.96	0.36	0.82	0.2	-34.6 to 35.0
OHIP-NL14	0.90	0.43	0.80	-0.0	-12.2 to12.1
OHIP-NL5	0.67	0.29	0.69	0.2	-0.5 to 0.8

Validity

Correlations between summary scores of the OHIP-NL, OHIP-NL14, and OHIP-NL5 were all high, and above 0.90 ($P < 0.001$), including the correlations between both short forms and the long form, when they were corrected for part-whole correlations, supporting the construct validity of the OHIPs.

Associations between the OHIP scores and pain-related disability scores were moderate, as shown in Table 2, supporting convergent validity for the original OHIP-NL as well as for the abbreviated OHIPs. Correlations between OHIP scores and oral health status were low, albeit significant ($P < 0.001$).

For group validity, associations between scores on all three OHIPs on the one hand and Characteristic Pain Intensity, the presence of burning mouth sensations, joint clicking, and biting activities on the other, are shown in Tables 3a (Spearman's rho) and 3b (T tests). The first two were moderate and significant ($P < 0.001$) and in the expected direction for all three OHIPs. As hypothesized, the associations between OHIP scores and joint clicking were not significant, but contrary to our expectations, biting activities were marginally, but significantly associated with the OHIP scores ($P < 0.05$).

Table 2. Convergent validity: Spearman's rho for the correlation between the mean summary scores of the OHIP-NL, the OHIP-NL14, and the OHIP-NL5 on the one hand, and tertiles of pain-related disability scores and the five levels of oral health status on the other hand

	N	OHIP-NL (mean±SD)	Association	OHIP-NL14 (mean±SD)	Association	OHIP-NL5 (mean±SD)	Association
Disability scores			0.46***		0.46***		0.39***
1. tertile	98	24.2±23.4		6.5±7.8		3.1±2.9	
2. tertile	65	33.0±25.2		9.6±8.3		4.4±3.1	
3. tertile	79	55.7±35.4		17.3±11.7		6.5±4.2	
Oral health status			0.28***		0.19***		0.21**
Excellent	18	34.9±27.7		10.6±9.5		4.0±3.2	
Very good	43	30.5±23.8		8.7±7.6		4.0±3.2	
Good	116	32.8±32.2		9.7±10.8		4.1±3.7	
Fair	56	47.3±32.4		13.7±11.1		5.8±3.9	
Poor	9	74.4±49.9		22.9±15.3		8.4±5.2	

*P < 0.05, ** P < 0.01, *** P < 0.001

Associations between scores on the OHIP-NL, the OHIP-NL14, and the OHIP-NL5 on the one hand, and the RDC/TMD axis II measures "jaw disability", "Somatisation", and "Depression" on the other hand, were moderate; those between OHIP scores and "Graded Chronic Pain" were low. They were all significant (P < 0.001) (see Table 4).

Combined, all these findings from the validity assessment were interpreted as evidence for the construct validity of all three OHIP versions.

Table 3a. Group validity: Spearman's rho for the correlation between the mean summary scores of the OHIP-NL, the OHIP-NL14, and the OHIP-NL5 on the one hand, and CPI (Characteristic Pain Intensity) and biting activities on the other hand

	N	OHIP-NL (mean ±SD)	Association	OHIP-NL14 (mean ±SD)	Association	OHIP-NL5 (mean ±SD)	Association
CPI			0.41***		0.42***		0.36***
1. tertile	85	22.7±24.3		6.1±8.2		2.8±3.0	
2. tertile	83	40.7±30.3		11.9±9.9		5.1±3.4	
3. tertile	72	50.6±36.9		15.7±12.1		6.1±4.4	
Biting activities			-0.13*		-0.15*		-0.14*
1. tertile	83	41.9±35.4		12.6±11.5		5.3±4.3	
2. tertile	93	39.5±33.2		11.6±11.2		4.8±3.7	
3. tertile	62	29.9±26.9		8.6±8.9		3.6±3.1	

*P < 0.05, ** P < 0.01, *** P < 0.001

Table 3b. Group validity: T-tests for associations between the mean summary scores of the OHIP-NL, the OHIP-NL14, and the OHIP-NL5 on the one hand, and burning mouth sensations and joint clicking on the other hand

	N	OHIP-NL (mean ±SD)	Association	OHIP-NL14 (mean ±SD)	Association	OHIP-NL5 (mean ±SD)	Association
Burning mouth sensations			0.34***		0.36***		0.35***
absent	198	33.2±28.0		9.5±9.2		4.1±3.3	
present	42	54.2±42.0		16.6±13.9		6.5±5.1	
Joint clicking			NS		NS		NS
absent	98	34.6±32.9		9.9±11.0		4.1±3.8	
present	145	39.4±32.4		11.8±10.6		5.0±3.8	

*P < 0.05, ** P < 0.01, *** P < 0.001

Table 4: Levels of impairment for the RDC/TMD axis II measures "GPC" (Graded Chronic Pain), "Jaw disability", "Somatisation", and "Depression": associations with the mean summary scores of the three OHIP versions (Spearman's rho)

	N	OHIP-NL (mean)	Spearman's rho	OHIP-NL14 (mean)	Spearman's rho	OHIP-NL5 (mean)	Spearman's rho
GCP			0.27***		0.26***		0.23***
0	60	32.8±36.0		9.6±12.0		4.0±4.1	
1	37	31.8±30.3		8.6±9.2		4.3±3.9	
2	79	35.8±21.5		10.8±7.7		4.4±2.6	
3	28	45.3±31.3		14.0±11.5		5.4±3.8	
4	20	68.1±52.1		20.1±16.4		8.1±5.8	
Jaw disability			0.34***		0.36***		0.41***
1. tertile	105	26.4±24.6		7.0±7.9		3.2±3.0	
2. tertile	94	42.9±33.7		13.3±11.2		5.2±3.9	
3. tertile	43	52.6±38.6		16.0±12.7		6.7±4.4	
Somatisation			0.36***		0.31***		0.27***
1. tertile	94	26.3±22.6		7.8±8.3		3.6±3.0	
2. tertile	80	36.1±32.7		10.5±10.2		4.4±3.6	
3. tertile	68	54.7±37.3		16.1±12.7		6.4±4.5	
Depression			0.42***		0.39***		0.30***
1. tertile	80	22.4±20.1		6.2±7.0		3.3±2.9	
2. tertile	83	35.4±28.8		10.5±9.7		4.3±3.3	
3. tertile	78	54.8±38.4		16.3±12.5		6.3±4.6	

*P < 0.05, ** P < 0.01, *** P < 0.001

Discussion

In this study, it was shown that the psychometric properties of the Dutch version of the OHIP-NL14 are as comparably good as those of the original version, the 49-item OHIP-NL. The test-retest reliability of the OHIP-NL5 was fair to good, as compared to excellent for the longer OHIPs. The internal consistency of the OHIP-NL5 did almost reach satisfactory levels (0.67, as compared to the required minimum of 0.70), but it did not meet the standards for clinical applications. This is not surprising, because while Cronbach's alpha might be artificially raised by adding more similar types of questions, it will be lowered when the amount of questions is drastically reduced. The findings regarding the OHIP-NL5 are very similar to those found for German TMD patients, where a Cronbach's alpha of 0.65 and an

inter-item correlation of 0.31 was reported (16). Other populations may have slightly higher values. For example, while scores in German general population subjects had a Cronbach's alpha of 0.76 and an average inter-item correlation of 0.39 (16), scores in Japanese prosthodontic patients even had a Cronbach's alpha of 0.81 (22). Thus, although the present results seem similar to the German ones (16), sampling variability, but also differences in setting, may have influenced the results. Hence, even though the very short OHIP-NL5 could be an attractive supplement in the OHIP-series, which could give a first crude overview of OHRQoL, on the basis of this study, the OHIP-NL14 is the preferred abbreviated OHIP version.

To compose the abbreviated OHIP versions, the items originally selected by Slade (15) and John et al. (16) were used. Although it would have been possible to select the items on the basis of a separate statistical analysis of the Dutch data, the method used in this study was preferred, because creating OHIP versions from the original instrument with different items for each translated version makes comparisons between studies difficult. This choice was supported by the fact that four different short OHIP versions, with different scoring methods, including the use of weighted items, were compared in a study by John et al. (16), in which it was shown that it hardly made any difference which version one used. The selection of a 14-item and a 5-item version was also made on pragmatic grounds. Reduction of items in a questionnaire unavoidably leads to some loss of information. To balance the desire to gather as much information as possible and the wish not to burden our patients with time-consuming paperwork, compromises have to be made. Given the psychometric properties of both abbreviated versions, the OHIP-14 seems to be the best compromise.

The validity measures of all three OHIP versions were highly significant and in the expected direction, even though most correlations were only moderate, according to Cohen's guidelines (31). The present results are comparable to those of John's study, who suggested that OHRQoL may be seen as a broad construct, which touches many other constructs. Therefore, the finding that the correlations with other constructs, in the present study, could not be qualified as high, was to be expected. OHRQoL reflects the personal impression of an individual about the impact of a complaint more directly than the oral complaints do. The results show that such impact is most clearly seen in a pain complaint: correlations with CPI, burning mouth sensations, and pain-related disability scores are highest. This is supported by a Dutch study using the OHIP-NL, comparing TMD patients with tooth wear patients and full denture patients, which showed that TMD patients scored significantly higher than the other two groups (33).

To determine the validity of the abbreviated OHIPs, the two factors determining the GCP, viz., pain intensity and the pain-related disability score, were used separately. The way in which they are combined in the GCP formula may lead to a loss of information, as was shown recently (34). In the original development of the GCP, the classification was created so as to mirror the medical classification of degrees of severity of an illness, and to create an easy to use, clear classification system (35). The hierarchical relationship between pain en disability was not perfect, however, and for example 10% of patients in grade III and 5 % of patients in grade IV did not have the assumed high pain intensity necessary for a Guttman scale (35). It is possible that the disability factor, which seems to be the most relevant

psychological measure in chronic TMD pain, is partly lost in this pain/disability formula. The reliability and validity of the separate factors was established in a recent study (36). Also in the present study, the two factors of the GCP were more highly correlated with OHQoL than the GCP itself. We therefore suggest changing the formula with which patients are classified with the GCP, and design a formula where the two factors are presented separately.

One of our validity variables behaved contrary to our expectation: "biting activities" turned out to be significantly correlated to OHRQoL, although the associations were low. This could be explained by the fact that, while these activities were irrelevant in a previously studied sample of partial and full denture wearers (14), for patients with TMD complaints, problems with the jaw and their impact may in some way directly or indirectly be related to oral parafunctions. In a recent study (37), it was shown that a majority of TMD patients consider oral parafunctions to be harmful to the jaw. Even though that study was mainly directed at the most frequent oral parafunction, namely bruxism, the outcome cannot be ignored. However, in several previous studies with TMD patients about the relationship between biting activities or joint clicking and OHRQoL, no significances were found (6; 12; 23), but in a Japanese study also small but significant correlations were found between OHRQoL and both biting and joint clicking (22). Both in our study and in the Japanese study, the correlations were much lower and only significant at the 0.05 level, showing a difference between a serious complaint like pain and a symptom that may generally be considered to have little impact, except in certain situations or populations. More studies, including the examination of cultural factors, are needed to assess this in more detail.

In the present study, both SCL90 measures (viz., somatisation and depression) were significantly associated with the OHIP scores. This result is similar compared to a study by John et al. (6), where statistically significant correlations were also found; however, the relationship between OHIP scores and somatisation was stronger than the relationship between OHIP scores and depression. It is probable that (similar to pain and pain-related disability), somatization and depression partly reflect the OHIP domains of psychological discomfort and psychological disability. Likewise, jaw disability may, at least in part, be reflected in the OHIP domain of functional disability.

As a measure of jaw disability, the original Jaw Disability Checklist (1) was used. The checklist was constructed by investigators working on the original RDC/TMD project team (1). Data are used on the basis of expert validity. It would have been preferable if we had data from the validated Jaw Function Limitation Scale (38), but these were not available for our study.

The significant correlations between RDC/TMD axis II measures and the different versions of the OHIP confirm the impression, suggested in the Introduction, that the axis II measures and the OHIP partly assess related constructs. These results are also comparable to other studies with TMD patients (6). Especially the high correlations between the OHIPs and the TMD pain-related disability score (a measure of the impact of TMD pain on daily, leisure, and work activities) and with jaw disability (a measure of the impact of TMD complaints on the functioning of the jaw) show, that the OHIP and these measures are sharing common ground. The correlations between OHIP scores and axis II variables were are all below 0.50, which means that each of these measures contributes to a part of the complaints-related

impact to which the other questionnaires do not contribute. It is not surprising that the OHIP, measuring seven different domains of OHRQoL, reflects not only the impact on functional problems (comparable to jaw disability) and on physical and psychological disability (comparable to the Graded Chronic Pain scale), but also the impact on psychological discomfort and handicap (39).

This study also has some limitations. The data of the abbreviated OHIPs were derived from the 49-item OHIP, and completing the long questionnaire may have influenced the results. For this reason, we suggest the study to be repeated with patients who only complete the 14 or 5 items. Another limitation of this study is, that no axis I diagnoses were available. In this study, we were mainly directed at the impact of TMD complaints. Judged from other TMD studies, we expect no large differences when the diagnosis is known. Nevertheless, it would be better if this would be evaluated in another study. It would also be informative if the TMD patients were then compared to a control group and possible floor effects were examined. A third limitation of the study is, that two variables, viz., joint clicking and burning mouth sensations, had dichotomous response options. It would have been preferable, if they had been formulated similar to the other variables. However, that information was not available. In our opinion, yes-no answers are less detailed, but still useful for showing associations between joint clicking or burning mouth sensations and their impact on quality of life. Furthermore, although some statistical analyses can show the difference between types of response scales, the outcomes of some studies using dichotomous response options versus Likert-type response scales have often yielded very similar results (26).

In the present study, short OHIP-NL versions were investigated in a consecutive sample of patients who were referred because of their TMD complaints. In our opinion, the short OHIP-NL14 scores have sufficient reliability and validity. More study is needed to examine whether the OHIP-NL5 can be used as a first global screening, when taking into account that the OHIP-NL5 has a marginal internal consistency, due to design of the instrument. Furthermore, even though the psychometric properties of the 49-item OHIP-NL were examined only with prosthodontic patients, the instrument has in the meantime also been used to compare TMD patients' OHRQoL with tooth wear and full denture patients. The study showed that the long Dutch OHIP version was be able to discriminate between TMD patients and tooth wear and full denture patient groups (33). Whether the results are generalizable to Dutch TMD patients in general, to dental patients, to general population subjects, or to other Dutch populations of interest cannot directly be inferred from our study. In our opinion, there are some arguments that the OHIP-NL14 can be used in other populations too: (1) the short form shows similar validity findings as the long form, and the OHIP-NL and OHIP-NL14 have been successfully used across settings and conditions; (2) other language OHIP short forms have been used successfully in other settings; (3) OHIP-NL14 has already been used in other settings in The Netherlands, but so far, no formal study has shown its psychometric properties; and (4) there are no convincing arguments not to generalize to other populations of interest. Validation is an ongoing process. However, even if this study has the limitation that our subjects cannot be characterized according to RDC/TMD axis I parameters, this situation probably does not compromise the direction and the magnitude of the relationships between OHRQoL measured by OHIP and other important

constructs. We may not be able to determine what the level of OHRQoL or, for example, the depression burden is in TMD patients (this requires a new study representative for TMD patients), but the relationship between these constructs that we found in our subjects should be relatively similar across settings.

Conclusions

The OHIP-NL and OHIP-NL14 both performed comparably well, and both better than the OHIP-NL5. When the length of the questionnaire (i.e., the time needed for its completion) is an issue, the OHIP-14 would therefore be the preferred questionnaire.

References

1. Dworkin SF, LeResche L. Research diagnostic criteria for temporomandibular disorders: review, criteria, examinations and specifications, critique. J Craniomandib Disord 1992; 6: 301-355.
2. Dworkin SF, Massoth DL. Temporomandibular disorders and chronic pain: disease or illness? J Prosthet Dent 1994; 72: 29-38.
3. Dahlström L & Carlsson GE. Temporomandibular disorders and oral health-related quality of life. A systematic review. Acta Odontol Scand 2010; 68: 80-85.
4. John MT, Koepsell TD, Hujoel P., Miglioretti DL., LeResche L , Micheelis W. Demographic factors, denture status and oral health-related quality of life. Community Dent Oral Epidemiol 2004; 32: 125-132.
5. John MT, Slade GD, Szentpétery A, Setz JM. Oral health-related quality of life in patients treated with fixed, removable, and complete dentures 1 month and 6 to 12 months after treatment. Int J Prosthodont 2004; 17: 503-511.
6. John MT, Reissmann D, Schiers O. Oral health-related quality of life in patients with Temporomandibular Disorders. J Orofac Pain 2007, 21: 46-54.
7. Ohrbach R., List T., Goulet J.-P. Svensson P. Recommendations from the International Consensus Workshop: convergence on an orofacial pain taxonomy. J Oral Rehabil 2010; 37: 807–812.
8. Slade GD, Spencer J. Development and evaluation of the Oral Health Impact Profile. Community Dental Health 1994; 11: 3-11.
9. John MT, Patrick DL, Slade PDL. The German version of the Oral Health Impact Profile – translation and psychometric properties. Eur J Oral Sci 2002; 110: 425-433.
10. Wong MCM, Lo ECM, McMillan AS. Validation of a Chinese version of the Oral Health Impact Profile (OHIP). Community Dent Oral Epidemiol 2002; 30: 423-431.
11. Larsson P, List Th, Lundström I, Marcusson A, Ohrbach R. Reliability and validity of a Swedish version of the Oral Health Impact Profile (OHIP-S). Acta Odontol Scand 2004; 62: 147-152.
12. Szentpétery A, Szabó G, Marada G, Szántó I, John MT. The Hungarian version of the Oral Health Impact Profile. Eur J Oral Sci 2006; 114: 197-203.
13. Bae KH, Kim C, Paik DI, Kim JB. A comparison of oral health related quality of life between complete and partial removable denture-wearing older adults in Korea. J Oral Rehabil 2006; 33: 317-322.
14. Meulen MJ van der, John MT, Naeije M, Lobbezoo F. The Dutch version of the Oral Health Impact Profile (OHIP-NL): translation, reliability and construct validity. BMC Oral Health 2008; 8: 11.
15. Slade GD. Derivation and validation of a short-form oral health impact profile. Community Dent Oral Epidemiol 1997; 25: 284-90.
16. John MT, Miglioretti DL, LeResche L, Koepsell TD, Hujoel P, Micheelis W. German short forms of the Oral Health Impact Profile. Community Dent Oral Epidemiol 2006; 34: 277-288.

17. Allen PF, Locker D. A modified short version of the oral health-related quality of life in edentulous adults. Int J Prosthodont 2002; 15 :446-450.
18. Locker D, Allen PF. Developing short-form measures of oral health-related quality of life. J Public Health Dent 2002; 62: 13-20.
19. Segu M, Collesano V, Lobbia S, Rezzani C. Cross-cultural validation of a short form of the Oral health Impact Profile for temporomandibular disorders. Community Dent Oral Epidemiol 2005; 33: 125-130.
20. Wong AHH, Cheung CS, McGrath C. Developing a short form of Oral Health Impact Profile (OHIP) for dental aesthetics: OHIP-aesthetic. Community Dent Oral Epidemiol 2007; 35: 64–72.
21. Fernandes MJ, Ruta DA, Ogden GR, Pitts NB, Ogston SA. Assessing oral health-related quality of life in general dental practice in Scotland: validation of the OHIP-14. Community Dent Oral Epidemiol 2006; 34:53-62.
22. Baba K, Inukai M, John MT. Feasibility of oral health-related quality of life assessment in prosthodontic patients using abbreviated Oral Health Impact Profile questionnaires. J Oral Rehabil 2008; 35: 224–228.
23. Baran I,Nalcaci R. Self reported problems before and after prosthodontic treatments according to newly created Turkish version of oral health impact profile. Arch Gerontol Geriatr 2010 (doi:10.1016/j.archger.2010.07.002).
24. Cronbach LJ. Coefficient Alpha and the internal structure of tests. Psychometrika 1951; 16: 297-334.
25. Bland JM, Altman DG. Cronbach's alpha. BMJ 1997; 314: 572.
26. Clark LA, Watson D. Constructing validity : Basic issues in objective scale development. Psychol Assess 1995; 7: 309-319.
27. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. Psychol Bull 1979; 86: 420-428.
28. Bland JM, Altman DG. Statistical methods of assessing agreement between two methods of clinical measurement. Lancet. 1986; 1: 307-311.
29. Fleiss JL. The design and analysis of clinical experiments. New York, NY: Wiley; 1986.
30. Cohen J, Cohen P, West SG, Aiken LS. Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences. Mahwah NJ: Lawrence Erlbaum Associates, 2003, 3rd ed.
31. Arrindell WA, Ettema JHM. Symptom Checklist SCL-90: Handleiding bij een Multi-dimensionele Psychopathologie-indicator. Lisse: Swets Test Publishers; 1986.
32. Cohen J. Statistical power analysis for the behavioral sciences. Hillsdale, NJ: Lawrence Erlbaum, 1988, 2nd ed.
33. Papagianni CE, Meulen MJ van der, Naeije M, Lobbezoo F. Orale Kinesiologie en levenskwaliteit. Ned Tijdschr Tandheelkd 2011; 118: 145-147.
34. Meulen MJ van der, Lobbezoo F, Aartman IHA, Naeije M. Ethnic background as a factor in temporomandibuar disorder complaints. J Orofac Pain 2009; 23: 38-47.
35. Von Korff M, Ormel J, Keefe FJ, Dworkin SF. Grading the severity of chronic pain. Pain 1992; 50: 133-149.

36. Ohrbach R, Turner JA, Sherman JJ, Mancl LA, Truelove EL, Schiffman EL, Dworkin SF. The Research Diagnostic Criteria for Temporomandibular Disorders. IV: evaluation of psychometric properties of the Axis II measures. J Orofac Pain 2010; 24: 48-62.
37. Meulen MJ van der, Ohrbach R, Aartman I, Naeije M, Lobbezoo F. TMD patients' illness beliefs and self-efficacy related to bruxism. J Orofac Pain 2010; 24:367-372.
38. Ohrbach R, Larsson P, List Th. The Jaw Functional Limitation Scale: development, reliability, and validity of 8-item and 20-item versions. J Orof Pain 2007; 22: 219-230.
39. Ohrbach R. Disability assessment in temporomandibular disorders and masticatory system rehabilitation. J Oral Rehab 2010; 37: 452-480.