Community Dent Oral Epidemiol 2008; 36: 12–20

COMMUNITY DENTISTRY AND ORAL EPIDEMIOLOGY

Effect of reducing the number of items of the Oral Health Impact Profile on responsiveness, validity and reliability in edentulous populations

Manal Awad[1], Muneera Al-Shamrany[2], David Locker[3], Finbarr Allen[4] and Jocelyne Feine[2]

[1]College of Dentistry, University of Sharjah, Sharjah, UAE, [2]Faculty of Dentistry, McGill University, Montred, Canada, [3]Faculty of Dentistry, University of Toronto, Toronto, Canada, [4]University Dental School and Hospital, National University of Ireland, Cork, Ireland

Awad M, Al-Shamrany M, Locker D, Allen F, Feine J. Effect of reducing the number of items of the Oral Health Impact Profile on responsiveness, validity and reliability in edentulous populations. Community Dent Oral Epidemiol 2008; 36: 12–20. © 2007 The Authors. Journal compilation © 2007 Blackwell Munksgaard

Abstract – *Background:* The 49-item Oral Health Impact Profile (OHIP) has shown strong responsiveness, reliability and validity. However, the large number of items included may limit its use in clinical trials, clinical practice and surveys. *Objective:* The main objective of this study is to assess the effect of reducing the number of items in each domain, one at a time, on responsiveness, reliability and validity of the OHIP in edentulous populations. *Materials and methods:* Data used in this study were obtained from two randomized clinical trials comparing mandibular implant overdentures and conventional dentures among 102 subjects between 35 and 65 years of age, and 60 subjects over the age of 65 years. Participants were edentulous individuals who wished to replace their current prostheses. Subjects in both trials were asked to complete the 49-item OHIP prior to treatment and at 2 months post-treatment. Within the study, effect sizes were computed at each stage of item reduction using the impact method. Intraclass correlation coefficients and Pearson's correlation coefficients were also assessed at each stage of item reduction. In addition, receiver-operating characteristic (ROC) curves were used to indicate the accuracy with which measurement changes corresponded to judgements of important changes in Oral Health Related Quality of Life (OHRQL). *Results:* The results indicated that, in general, domain responsiveness was not affected by the reduction of the number of items used per domain. However, there was a decrease in reliability, especially within the 'psychological' and 'social' disabilities and 'handicap' domains (35- to 65-year group). In addition, there was a decrease in construct validity of the 'physical pain', 'psychological' and 'social disabilities' domains (35- to 65-year group), as well as on 'physical pain', 'psychological discomfort', 'physical' and 'psychological' disabilities in the 65-year and older group. This occurred primarily, when reducing from two to one item per domain. Among the 35- to 65-year group, there were consistencies in patients' ratings of the importance of similarly measured changes in oral health. *Conclusion:* The results indicate that although the 49-item OHIP responsiveness could be maintained with item reduction, this will lead to compromises in reliability and validity.

Key words: edentulism; oral health-related quality of life; reliability; responsiveness; validity

Dr Manal Awad, College of Dentistry, University of Sharjah, Sharjah PO Box 27272, United Arab Emirates
Tel: +971-6-5053184
Fax: +971-6-5050802
e-mail:awad@sharjah.ac.ae

Submitted 14 March 2006;
accepted 24 July 2006

There is an increased recognition of the importance of incorporation of Oral Health Related Quality of Life (OHRQL) measurements in evaluations of oral health (1–3). Assessment of OHRQL in clinical trials assists in providing a better understanding of treatment outcomes from the patient's perspective.

doi: 10.1111/j.1600-0528.2006.00364.x

In addition, OHRQL instruments could be useful for clinicians to monitor a patient's condition in clinical practice. However, one major obstacle in using such instruments is the time required for completion by patients particularly in a clinical trial setting, where patients are usually given several questionnaires to complete. In addition, clinicians may find it time consuming and difficult to interpret the information obtained from long assessment instruments (4). Therefore, it is more efficient to use shorter questionnaires that capture important aspects of change in patients' oral health status. Nevertheless, shortening original instruments could affect its precision and psychometric properties such as responsiveness, validity and reliability.

The Oral Health Impact Profile (OHIP) is an instrument that was specifically designed to measure the impact of oral health on quality of life (5). This instrument consists of seven domains (functional limitation, physical pain, psychological discomfort, physical disability, psychological disability, social disability and handicap). The reliability, validity and responsiveness of this instrument have been previously established (5, 6). In addition, it has been used in clinical trials to evaluate the effectiveness of treatment for edentulism (1–3). Reports from these studies imply that this instrument discriminates between edentulous subjects according to treatment received (1) and demonstrates sensitivity to pre-/post-treatment change in OHRQL (2, 3).

As the OHIP is a relatively long questionnaire, several suggestions for shorter versions were published (7–9). A first short version was introduced by the developer of the original 49-item questionnaire (7). Using data from a cross-sectional study of 1217 Australian dentate and edentulous individuals, Slade (7) used three statistical methods to reduce the questionnaire's length, namely, principal component factor analysis, least square regression and internal reliability analysis. A 14-item short-form OHIP was derived based on the least square method. Using the impact method, Allen and Locker (8) proposed an alternative 14-item short form that had only two items in common with the regression method previously developed. The authors suggested that if the main aim is to detect change in OHRQL over time, the impact method maybe the preferable method of shortening questionnaires. In a subsequent publication, the same authors (9) developed another 19-item OHIP questionnaire (OHIP-EDENT) specifically for an edentulous population. These results showed that the 19-item short form had acceptable discriminant validity and was able to detect change in edentulous patients' ratings of their new prostheses.

While there are no clear rules about by how much one can reduce an original questionnaire, some researchers have asserted that, for a particular domain to be measured appropriately, one needs to use more than two items per domain, which will reduce the possible effect of eccentric responses to individual questions (10). In addition, it is possible that reducing the original questionnaire by more than 50% of its original length may affect the instrument's construct validity (11). Moreover, reduction of items of a health-related quality of life instrument could lead to the omission of individual patient problems and a loss of content validity.

Therefore, data from two Canadian randomized clinical trials among edentulous populations, comparing mandibular two-implant overdentures and conventional dentures were used to assess at what point the deletion of items of the original OHIP begins to compromise the performance of the instrument.

Materials and methods

Selection and description of participants

The first randomized control trial (RCT) was conducted to compare mandibular two-implant overdentures and conventional dentures among 102 edentulous individuals between the ages of 35 and 65 years (1) and the second RCT was conducted to compare the same types of prostheses among 60 edentulous subjects over the age of 65 years (2). In both trials, participants were asked to respond to the 49-item OHIP prior to treatment and at 2 months post-treatment.

Technical information

Originally, the OHIP responses comprised five categories (5); however, in both RCTs the response categories were extended to include an additional response, to allow for the broadest number of response categories that patients can choose from (1, 2). Therefore, subjects were asked to indicate the frequency of occurrence of a particular problem by selecting from the following categories: 'never', 'rarely', 'occasionally', 'some of the time', 'most of the time' and 'all of the time'. Coding ranged from '1' to '6' respectively. In both trials at 2 months

post-treatment, those who received mandibular two-implant overdentures had significantly fewer health-related quality of life problems than those who received conventional dentures (1, 2).

Statistical analysis

Item impact

Several methods are available for item reduction. These methods rely on statistical procedures, such as regression modelling and factor analysis (7, 11), or on patients' perceptions of level of importance of different items, such as the item impact method (8, 9). It has been suggested that a disease-specific quality of life measurement should include items that are primarily important to the patients, irrespective of their association with each other (12). Using this method requires dichotomizing the responses according to whether or not an important change occurred in the patients' OHRQL, as a result of the intervention. Therefore, in this study, the impact method was used to reduce items by calculating the proportion of subjects with responses 'some of the time' coded as '4', to 'all of the time' coded as '6'. In addition, the mean frequency ratings of subjects with these responses were also obtained. This mean frequency rating was calculated by summing the responses coded as 'some of the time' to 'all of the time' and dividing this sum by the number of subjects with those responses. An item impact was obtained by multiplying these two values (mean frequency and the proportion of subjects with these responses) and the items' weights, which were previously calculated in a Canadian population (13).

Item responsiveness

There is a consensus in the literature that responsiveness is the instrument's ability to detect change. Specifically, it was defined as 'the ability to detect meaningful treatment effects' (14). However, there are a variety of opinions regarding the nature of the change being detected (15–19). One method of assessment of change is to include data only from subjects who exhibited improvement (15). The data used in this study were obtained from two clinical trials specifically designed to measure patients' oral health status with implant-supported prostheses compared with new conventional dentures. Therefore, we opted to include data only from subjects deemed to have shown improvement, according to a valid external criterion that represents the 'gold standard'. Several external criteria have been used to classify change as clinically important. Previous studies have used a variety of outcomes for this purpose, such as, satisfaction with care (18, 19), pain improvement or change in health status and return to normal activities (17, 20). In this study, we used satisfaction with the new prostheses as the gold standard, and this was the primary outcome in both studies. Therefore, for the purpose of this study, improvement was defined as a positive change of ≥20 mm on the rating of satisfaction with the new prostheses using 100-mm visual analogue scales (VAS).

Item responsiveness among improved subjects was computed using effect sizes on the differences in OHIP mean values between pre- and post-treatment scores divided by the standard deviation of the difference (8, 21). Effect sizes were computed for all items in each domain of the OHIP using data from each of the two studies.

An initial step in reducing the instrument was removal of items that correlated highly with another item (0.7) and combining them into one (20). Based on the impact method items were removed, one at a time, beginning with the least responsive. At each stage of reduction a new domain score was computed. These procedures were performed on both studies.

Assessing measurement properties

In addition to calculating responsiveness, validity and reliability were also assessed for each domain score at each step of item reduction.

Reliability

During two consecutive weeks, and prior to initiation of the new treatments, subjects were asked to complete questionnaires regarding their satisfaction with their current prostheses. Intra-class correlation coefficients (ICC) for test–retest reliability was computed using the variance between patients divided by the sum of the variances between and within patients. This property was assessed in the 35- to 65-year group only, as data were not available for the over 65-year population.

Validity

Construct validity was assessed in both studies using correlations between change scores measured for each domain at each stage of item reduction and change in the global rating of general satisfaction.

Receiver-operating characteristic (ROC) curves were used to indicate the accuracy with which measurement changes corresponded to judgements

of important changes in OHRQL, using post-treatment change in satisfaction with prostheses as the gold standard (patients who reported significant improvement versus patients who did not report significant improvement). An area under the ROC curve of 0.5 would indicate that, overall the magnitudes of change in OHRQL measures were no better at identifying an important change than a random guess (22). This result would be obtained if there was little consistency among patients in the magnitude of change they consider important. An area under the ROC curve that was significantly greater than 0.5 would indicate that, measured changes in the patient global assessment were associated with patients' judgements that a change of any degree of importance had occurred. It also indicates some consistency in these judgements among patients. In addition, if the OHIP questionnaire is considered a diagnostic test for improvement, then this instrument can also be described in terms of sensitivity and specificity for detecting change as established by the gold standard (23). The cut-off point of measured change that had the highest sensitivity and specificity was also identified.

Results

As an initial step, the following items were removed: item no. 3 (Noticed tooth that doesn't look right), no. 13 (Sensitive teeth), and no. 14 (Toothache) and no. 27 (Unable to brush teeth), as they did not apply to edentulous populations and the subjects did not respond to them. In addition, redundancy was removed by combining items that had item–item correlation that was 0.7 or above in both studies. The following pairs: items no. 1 (Difficulty chewing) and no. 16 (Uncomfortable to eat) no. 10 (Painful aching) and no. 17 (Sore spots), no. 16 and no. 28 (Avoid eating) and no. 21 (Miserable) and no. 23 (Tense) were highly correlated, above 0.7. The expert panel viewed these items for redundancy and decided to remove items 10, 16 and 21. The instrument at this stage included 42 items.

The analyses of effect sizes at each stage of item removal were conducted among patients who reported significant improvement in ratings of general satisfaction [63 (62%) subjects in the 35- to 65-year group and 42 (70%) in the over 65-year group]. However, the sociodemographic characteristics of these subjects (age, gender, level of education and marital status), as well as satisfaction with initial prostheses were not different from those who did not report significant improvement with received prostheses. Item reduction started with the items with the lowest impact, retaining the item with the greatest impact in each domain. With the exception of 'functional limitation' and 'social disability' domains in the 35- to 65-year group and 'physical pain' and 'handicap' domains in the over 65-year group there was a slight increase in responsiveness with one item per domain compared with the original 49-item OHIP (Table 1).

The computed effect sizes in both groups were equally high, with the 'functional limitation' domain having the greatest responsiveness at all stages of item removal. This was consistently observed for both study populations. In four of the seven domains, the same item was retained for both groups, namely: 'difficulty chewing', 'tense', 'less tolerant of others' and 'life unsatisfactory'. For the remaining three domains there was no great discrepancy in the order of item removal. For example, in the 'physical disability' domain, 'avoid eating' was the item to be retained in the 35- to 65-year group, while in the older group this item was removed last and the retained item was 'speech unclear' (Table 1).

Table 2 depicts the results of the test–retest reliability using ICC in the 35- to 65-year group. In general, there was a very small decrease in the reliability of each domain with the reduction of items. However, a more noticeable decrease was observed with the reduction from two to one item per domain. The greatest decrease occurred in the 'functional' and psychological discomfort' domains (10% and 12%, respectively). Even with the reduction from the original to two items per domain scale, the reduction in reliability ranged from 13% for social disability domain to as low as 1% in the physical pain domain.

Similarly, there was a decrease in correlation coefficients with the reduction of items between pre- and post-change scores of the OHIP and global ratings of change in general satisfaction (Tables 3 and 4). This decrease in construct validity was marked when domains were reduced from two to one item in the 'physical pain domain' (18%) in the 35- to 65-year group and physical pain and physical disability domains (12%) in the over 65-year-old group.

Among the 35- to 65-year group, the areas under the ROC curves were all significantly greater than 0.5. An increase in the area under the ROC curve

Table 1. Effect sizes of the Oral Health Impact Profile (OHIP) in both 35- to 65-year and >65-year groups according to order of item removal, using item impact method

	35–65 years (n = 63)		Over 65 years (n = 42)	
Domains	Effect sizes	Order of removal	Effect sizes	Order of removal
Functional limitation				
Q1 Difficulty chewing	2.01	Retained	2.17	Retained
Q2 Trouble pronouncing words	2.01	3	1.87	3
Q4 Appearance affected	2.16	4	2.15	4
Q5 Breath stale	2.00	2	2.01	2
Q6 Taste worse	2.01	1	2.02	1
Q7 Food catching	2.36	6	1.75	5
Q8 Digestion worse	2.35	5	2.05	6
Q9 Dentures not fitting	2.10	7	2.05	7
Physical pain				
Q11 Sore jaw	1.77	2	2.02	4
Q12 Headaches	1.66	1	1.40	1
Q15 Painful gums	1.85	Retained	1.57	3
Q17 Sore spots	1.76	3	1.64	Retained
Q18 Discomfort (dentures)	1.90	4	1.68	2
Psychological discomfort				
Q19 Worried	1.53	3	1.54	2
Q20 Self-conscious	1.55	1	1.60	1
Q22 Appearance	1.49	2	1.66	3
Q23 Tense	1.90	Retained	1.76	Retained
Physical disability				
Q24 Speech unclear	2.30	7	1.89	Retained
Q25 Others misunderstood	1.76	4	1.82	5
Q26 Less flavor in food	1.70	2	1.63	2
Q28 Avoid eating	2.30	Retained	1.74	6
Q29 Diet unsatisfactory	1.68	3	1.66	3
Q30 Unable to eat (denture)	1.75	5	1.77	7
Q31 Avoid smiling	1.64	1	1.73	1
Q32 Interrupted meals	1.96	6	1.72	4
Psychological disability				
Q33 Sleep interrupted	1.11	1	1.71	1
Q34 Upset	1.20	2	1.67	3
Q35 Difficult to relax	1.70	Retained	1.68	5
Q36 Depressed	1.30	4	1.89	Retained
Q37 Concentration	1.22	3	1.67	4
Q38 Been embarrassed	1.26	5	1.70	2
Social disability				
Q39 Avoid going out	1.15	1	2.15	4
Q40 Less tolerant of others	2.1	Retained	2.13	Retained
Q41 Trouble getting on with others	1.10	2	1.54	2
Q42 Irritable with others	2.14	4	2.15	3
Q43 Difficulty doing jobs	2.0	3	1.67	1
Handicap				
Q44 Health worsened	1.30	4	1.89	5
Q45 Financial loss	1.36	1	1.75	1
Q46 Unable to enjoy people's company	1.35	2	1.98	4
Q47 Life unsatisfying	1.78	Retained	1.85	Retained
Q48 Unable to function	1.49	5	2.01	3
Q49 Unable to work	1.29	3	1.98	2

was observed at different stages of reduction of the functional limitation, physical pain, physical disability and psychological disability domains (Table 5). For example, in the functional limitation scale, the area under the ROC curve was 0.68 (95% CI 0.57–0.80), whilst with one item retained, the area under the ROC curve was 0.73 (95% CI 0.60–0.84). Increase in scale sensitivity was observed with item reduction in the physical and psychological disability scales. However, with other scales, such as social disability and handicap, almost no change was observed in sensitivity. Among the

Table 2. Test–retest reliability[a] for the two pretreatment visits at each stage of item removal among the 35- to 65-year group ($n = 102$).

Domains	Original	7	6	5	4	3	2	1
Functional limitation	0.56	0.60	0.59	0.60	0.61	0.59	0.60	0.50
Physical pain	0.61				0.61	0.60	0.63	0.62
Psychological discomfort	0.59					0.59	0.60	0.48
Physical disability	0.65	0.59	0.60	0.61	0.60	0.60	0.61	0.56
Psychological disability	0.72			0.70	0.70	0.70	0.61	0.62
Social disability	0.68				0.64	0.55	0.48	0.53
Handicap	0.72			0.68	0.67	0.65	0.62	0.57

[a]Reliability was measured using the intraclass correlation coefficient (ICC).

Table 3. Correlations between post-treatment change in OHIP scores and global change in ratings of general satisfaction at each stage of item reduction among the 35- to 65-year group ($n = 63$)

Domains	Original	7	6	5	4	3	2	1
Functional limitation	0.76**	0.78**	0.79**	0.82**	0.82**	0.82**	0.82**	0.77**
Physical pain	0.78**				0.73**	0.77**	0.78**	0.60**
Psychological discomfort	0.67**					0.67**	0.64**	0.61**
Physical disability	0.69**	0.69**	0.67**	0.64**	0.66**	0.65**	0.63**	0.61**
Psychological disability	0.60**			0.60**	0.61**	0.61**	0.53**	0.49*
Social disability	0.50*				0.50*	0.50*	0.50*	0.41*
Handicap	0.52*			0.53**	0.53**	0.54**	0.53**	0.50*

*$P \leq 0.05$; **$P \leq 0.01$.

Table 4. Correlations between post-treatment change in OHIP scores and global change in ratings of general satisfaction at each stage of item reduction among the >65-year group ($n = 42$)

Domains	Original	7	6	5	4	3	2	1
Functional limitation	0.33*	0.35*	0.37*	0.40*	0.38*	0.42*	0.51*	0.52*
Physical pain	0.36*				0.33	0.25	0.30*	0.18
Psychological discomfort	0.49*					0.44	0.36*	0.25
Physical disability	0.41*	0.35*	0.41*	0.37*	0.39*	0.42*	0.40*	0.28
Psychological disability	0.46*			0.44*	0.43*	0.41*	0.47*	0.41*
Social disability	0.25				0.24	0.18	0.21	0.23
Handicap	0.14			0.12	0.12	0.11	0.15	0.08

*$P \leq 0.05$.

over 65-year group, areas under the ROC curves, all 0.5 or less, were not significant.

Discussion

There is a lack of clarity in relation to how many items in each conceptual domain of the OHIP are required for maintaining the psychometric properties of this instrument at a practical, but also acceptable level.

In general, the results of this study show that responsiveness was not affected by a reduction in the number of items in each domain; for the most part, this was observed in both studies. However, for some domains a reduction of items appears to improve responsiveness, which was reflected in an increase in effect size. These findings were also reported by Moran et al. (24), who demonstrated that shortening the 20-item Chronic Respiratory Questionnaire (CRQ) led to improvement in the responsiveness in eight of 12 domains. Their findings were consistent across data from three different studies. Notably, the reported effect sizes from this study are all considered high. This is attributed to the fact that change scores were considered only for those who reported post-treatment improvement, and that patients in both studies were initially dissatisfied with their current prostheses and wished to replace it.

The results of item reduction in this study are generally in agreement with those previously

Table 5. Receiver-operating characteristics (ROC) curve areas and cutoff points with sensitivity, specificity, positive predictive values (PPV) and Negative predictive values (NPV) for improvement in patients OHIP assessments among 35- to 65-year group ($n = 102$)

Domains	Original	7	6	5	4	3	2	1
Functional limitation								
Area under the curve (95% CI)[a]	0.68 (0.57–0.80)	0.68 (0.55–0.79)	0.67 (0.55–0.79)	0.69 (0.57–0.80)	0.70 (0.59–0.82)	0.71 (0.60–0.82)	0.71 (0.60–0.82)	0.73 (0.60–0.84)
Cut-off point	16	14	13	10	9	8	5	2
Sensitivity/specificity	0.85/0.57	0.77/0.60	0.79/0.59	0.66/0.61	0.77/0.61	0.84/0.58	0.85/0.54	0.77/0.61
PPV/NPV	0.69/0.66	0.74/0.64	0.74/0.65	0.71/0.61	0.75/0.61	0.68/0.63	0.73/0.71	0.75/0.64
Physical pain								
Area under the curve	0.73 (0.63–0.84)				0.74 (0.63–0.84)	0.75 (0.65–0.84)	0.77 (0.67–0.87)	0.73 (0.63–0.84)
Cut-off point	10				9	8	5	2
Sensitivity/specificity	0.77/0.54				0.77/0.54	0.72/0.61	0.79/0.56	0.67/0.66
PPV/NPV	0.71/0.61				0.71/0.62	0.73/0.60	0.73/0.64	0.74/0.56
Psychological discomfort								
Area under the curve	0.69 (0.57–0.80)					0.69 (0.58–0.80)	0.71 (0.60–0.81)	0.67 (0.56–0.78)
Cut-off point	6					4	2	1
Sensitivity/specificity	0.80/0.56					0.72/0.53	0.71/0.66	0.84/0.49
PPV/NPV	0.72/0.65					0.69/0.55	0.75/0.60	0.71/0.67
Physical disability								
Area under the curve	0.67 (0.56–0.78)	0.67 (0.56–0.78)	0.66 (0.55–0.78)	0.66 (0.55–0.78)	0.67 (0.57–0.79)	0.67 (0.56–0.78)	0.68 (0.57–0.80)	0.69 (0.58–0.80)
Cut-off point	13	11	10	9	8	7	5	2
Sensitivity/specificity	0.71/0.53	0.64/0.59	0.69/0.57	0.69/0.57	0.79/0.54	0.74/0.52	0.77/0.57	0.74/0.54
PPV/NPV	0.68/0.54	0.70/0.52	0.70/0.55	0.70/0.55	0.72/0.63	0.69/0.57	0.67/0.56	0.70/0.58
Psychological disability								
Area under the curve	0.69 (0.59–0.80)			0.70 (0.59–0.81)	0.70 (0.59–0.80)	0.70 (0.59–0.81)	0.70 (0.60–0.81)	0.62 (0.51–0.74)
Cut-off point	7			6	5	3	3	2
Sensitivity/specificity	0.71/0.64			0.72/0.54	0.77/0.53	0.66/0.66	0.65/0.66	0.82/0.40
PPV/NPV	0.69/0.55			0.69/0.56	0.71/0.61	0.76/0.77	0.75/0.77	0.68/0.61
Social disability								
Area under the curve	0.63 (0.52–0.75)				0.64 (0.53–0.75)	0.64 (0.53–0.75)	0.62 (0.50–0.73)	Not significant
Cut-off point	5				4	3	2	
Sensitivity/specificity	0.82/0.44				0.84/0.44	0.84/0.45	0.85/0.38	
PPV/NPV	0.68/0.62				0.69/0.64	0.69/0.63	0.67/0.63	
Handicap								
Area under the curve	0.65 (0.54–0.76)			0.64 (0.53–0.76)	0.64 (0.53–0.76)	0.64 (0.52–0.75)	0.63 (0.52–0.75)	0.64 (0.52–0.75)
Cut-off point	7			5	4	3	2	1
Sensitivity/specificity	0.82/0.48			0.82/0.44	0.82/0.44	0.82/0.43	0.85/0.39	0.85/0.39
PPV/NPV	0.69/0.63			0.68/0.62	0.67/0.63	0.68/0.62	0.68/0.64	0.68/0.64

[a]95% confidence intervals.

reported from a Canadian study of older adults and with a longitudinal British study comparing implant supported overdentures and conventional dentures among patients who requested and received their preferred treatment (8). For example, in the 'functional limitation' domain, the two items retained by the Canadian group were 'difficulty chewing' and 'food catching', while in the British group, 'dentures not fitting' replaced the latter. In this study these three items had the highest impacts and were removed in the last stages.

Compared with the original OHIP, there was a decrease in reliability and construct validity when the number of domains were reduced. Specifically, test–retest reliability was affected when reducing from two to one item per domain. In addition, in two domains (handicap and psychological disability) the drop in the correlation coefficient occurred when the domains were reduced from three items to two. This observation indicates that the stability of the instrument could be compromised, especially when using only one item per domain.

In this study, it was demonstrated that among the 35- to 65-year group, there were consistencies in patients' ratings of the importance of similarly measured changes in oral health. In all test variations, the areas under the ROC curves were significantly greater than 0.5, indicating that patients were consistent in their judgements of the magnitude of change in original and reduced OHIP. At all stages of item reduction, specificity was much lower than sensitivity. However, if sensitivity (the ability of the OHIP to correctly identify edentulous subjects who are satisfied with their oral prostheses) is valued more than specificity (the ability of the OHIP to correctly identify edentulous subjects who are not satisfied with their oral prostheses) for the identification of important change, then cut-off points obtained from ROC curves could be used to identify criterion standards that meet predetermined levels of sensitivity.

The above-mentioned associations were not observed among the over 65-year group, suggesting that older edentulous patients' ratings of satisfaction with oral prostheses may not necessarily reflect important change in OHRQL. In both RCTs there was a marked reduction in the correlation coefficient for the association between change in OHIP scores and global change in ratings of general satisfaction. This could indicate that content validity may be compromised and the comprehensiveness of the instrument could be jeopardized when a shorter version is used. Hence, the remaining items in a shorter version may not capture the complete picture of the patient's experience (11, 24). One possible solution is to ask patients to assess comprehensiveness of the instrument to maintain adequate content validity (24).

Previously developed short versions of the OHIP had been based on the notion that one needs to select an equal number of items per domain, even if different domains are represented by an unequal number of items (7–9). For example, in the original OHIP the 'functional limitation' (nine items), 'physical pain' (nine items), and 'physical disability' (eight items) domains include more items than the other four scales. This is reasonable, as these domains directly assess the impact of various oral health problems such as loss of teeth and other oral conditions on general oral functioning. Therefore, selection of a shorter version may take into consideration the weight given to a particular domain, given the number of items used to comprehensively measure an attribute. This may indeed enhance the content validity and reduce the risk of compromising the underlying construct on which the original questionnaire was built.

Using two different studies enhances the generalizability of the results, especially because the same treatment regimen was applied to both studies conducted among French Canadians. However, patients' assessments did not occur at the same time, and the difference in age between the two groups could explain the differences in the results of the two studies. For example, using item responsiveness is an attractive method to reduce OHIP domains, as it resulted in retaining items with the highest level of responsiveness (8, 11, 13, 20). However, in three of the seven OHIP domains a different item was retained from the two samples. Accordingly, it is conceivable that the reduction of items is data-dependent or a chance phenomenon and variations in retained items could be observed with other clinical samples (11, 24). If this is the case, then the same assumptions could apply to the development of any short form of an original questionnaire, and careful consideration must be given to the purpose of producing a shorter version of an instrument (9, 11, 24). Furthermore, results from this study should be interpreted with the understanding that the data used were obtained from edentulous populations only, and that four items were initially removed, as they did not apply to these populations.

In this study, as well as others (8, 9), the assessments of the reliability, validity and responsiveness of a shorter version of a QoL instrument were

obtained from patients' responses to the original questionnaires. Methods chosen for shortening assessments are made under the assumption that subjects' responses to individual items are not influenced by the context. However, it is possible that patients' responses to individual items may vary based on the number of items per domain (24). It would be informative if in future trials short and original OHIP questionnaires are given to patients on two consecutive occasions in order to test the hypothesis that item responses do not vary according to the number of items embedded in one domain.

In summary, the results of this study show that care must be given to the possible changes that could occur as a result of shortening an original instrument. Failure to address these issues in a scientific rather than an empirical fashion could produce measures not suited to the task for which they are being used.

Acknowledgements

These studies were generously supported by two CIHR Industry Grants with industrial partners Nobel Biocare Canada and Straumann Canada Limited.

References

1. Awad M, Locker D, Korner-Bitensky N, Feine J. Measuring the effect of implant rehabilitation on health related quality of life in a randomized clinical trial. J Dent Res 2000;79:1659–63.
2. Awad MA, Lund JP, Shapiro SH, Locker D, Klemetti E, Chehade A et al. Oral health status and treatment satisfaction with mandibular implant overdentures and conventional dentures: a randomized clinical trial in a senior population. Int J Prosthodont 2003;16:390–6.
3. Allen PF, McMillan AS. The impact of tooth loss in a denture-wearing population: an assessment using the Oral Health Impact Profile. Community Dent Health 1999;16:176–80.
4. Jones G, Jenkinson C, Kenndey S. Development of the short Endometriosis Health Profile Questionnaire: The EHP-5. Qual Life Res 2004;13:695–704.
5. Slade G, Spencer A. Development and evaluation of the Oral Health Impact Profile. Community Dent Health 1994;11:3–11.
6. Slade GD. Assessing change in quality of life using the Oral Health Impact profile. Community Dent Oral Epidemiol 1998;26:52–61.
7. Slade G. Derivation and validation of a short-form oral health impact profile. Community Dent Oral Epidemiol 1997;25:284–90.
8. Allen F, Locker D. A modified short version of the Oral Health Impact profile for assessing health related quality of life in edentulous adults. Int J Prosthdont 2002;15:446–50.
9. Locker D, Allen FP. Developing short-form measures of oral health related quality of life. J Public Health Dent 2002;62:13–20.
10. Guyatt GH, Kirshner B, Jaeschke R. Measuring health status: What are the necessary measurement properties? J Clin Epidemiol 1992;12:1341–5.
11. Coste J, Guillemin F, Pouchot J, Femanian J. Methodological approaches to shortening composite measurement scales. J Clin Epidemiol 1997;50:247–52.
12. Juniper EF, Guyatt GH, Streiner DL, King DR. Clinical impact versus factor analysis for quality of life questionnaire construction. J Clin Epidemiol 1997;50:233–8.
13. Allison P, Locker D, Jokovic A, Slade D. A cross cultural study of oral health values. J Dent Res 1999;78:643–9.
14. Testa M, Nackley J. Methods for quality of life studies. Annu Rev Public Health 1994;15:535–59.
15. Beaton DE, Bombardier C, Katz JN, Wright JG. A taxonomy for responsiveness. J Clin Epidemiol 2001;54:1204–17.
16. Deyo RA, Inui TS. Towards clinical applications of health status measures: sensitivity of scales to clinical important changes. Health Serv Res 1984;19:275–89.
17. Beaton DE, Hogg-Johnson S, Bombardier C. Evaluating changes in health status: reliability and responsiveness of five generic health status measures in workers with musculoskeletal disorders. J Clin Epidemiol 1997;50:79–93.
18. Stucki G, Liang MH, Fossel AH, Katz JN. Relative responsiveness of condition-specific and generic health status measures in degenerative lumbar spinal stenosis. J Clin Epidemiol 1995;48:1369–78.
19. Sullivan M, Karlsson J. The Swedish SF-36 Health Survey III. Evaluation of criterion-based validity:results from normative population. J Clin Epidemiol 1998;11:1105–13.
20. Juniper EF, Guyatt GH, Cox FM, Ferrie PJ, King DR. Development and validation of the Mini Asthma quality questionnaire. Eur Respir J 1999;14:32–8.
21. Bronfort G, Bouter LM. Responsiveness of general health status in chronic low back pain: a comparison of the COOP charts and the SF-36. Pain 1999;83:201–9.
22. Hanley JA, McNeil BJ. Yje. The meaning and use of the area under a receiving operating characteristic (ROC) curve. Radiology 1982;143:29–36.
23. Deyo RA, Centor RM. Assessing responsiveness of functional scales to clinical change; an analogy to diagnostic test performance. J Chronic Dis 1986; 39:897–906.
24. Moran LA, Guyatt GH, Norman GR. Establishing the minimal number of items for a responsive, valid, health-related quality of life instrument. J Clin Epidemiol 2001;54:571–9.