Community Dent Oral Epidemiol 2007; 35: 64–72

COMMUNITY DENTISTRY AND ORAL EPIDEMIOLOGY

Developing a short form of Oral Health Impact Profile (OHIP) for dental aesthetics: OHIP-aesthetic

Anthony H. H. Wong, C. S. Cheung and Colman McGrath

Periodontology & Public Health, Faculty of Dentistry, University of Hong Kong, Hong Kong SAR, China

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. Journal compilation © 2007 Blackwell Munksgaard

Abstract – *Objectives:* To develop and evaluate shortened forms of the Oral Health Impact Profile (OHIP) for discriminating dental aesthetics problems and evaluating dental aesthetic outcomes. *Methods:* Eighty-seven subjects self-completed the 49-item OHIP at baseline and 63 at follow up (8 weeks later), with the intervention of applying an array of tooth-whitening products. Expert-based approach and regression analysis (on baseline data) were undertaken to derive two subset questionnaires (OHIP-conceptual and OHIP-regression). Their discriminatory ability for dental aesthetics and their responsiveness to tooth whitening were compared with the original OHIP-49, Slade's OHIP-14 and a Chinese short-form version of OHIP. *Results:* The measures developed were strongly associated with self-rating of dental aesthetics ($P < 0.001$) unlike OHIP-49 ($P = 0.03$) or other OHIP short forms ($P > 0.05$). The measures were also reliable (Cronbach's alpha 0.86) and comparable with the other OHIP forms. In terms of effect size, OHIP-conceptual was more effective in measuring changes than the one based on the regression analysis, the original OHIP-49, OHIP-14 and the Chinese version of the short-form OHIP. It also exhibited a less susceptibility to floor effects than other OHIP forms. *Conclusion:* A modified short form of the OHIP derived (OHIP-conceptual) was the most favorable in discriminating dental aesthetics, was reliable and most sensitive to the dental aesthetics intervention – tooth whitening.

Key words: aesthetic dentistry; outcome measure; quality of life; tooth whitening

Dr Colman McGrath, Periodontology & Public Health, Faculty of Dentistry, University of Hong Kong, 34 Hospital Road, Hong Kong SAR, China
e-mail: mcgrathc@hkucc.hku.hk

Submitted 11 March 2005;
accepted 19 September2005

Dental aesthetics (namely tooth colour) is a key concern of the public, and is the reason why there are so many different forms of tooth-whitening products available in the market (1–3). Aesthetics is very much a subjective perception that varies from individual to individual and therefore it is difficult to assess dental aesthetics or evaluate the effectiveness of any intervention aimed at altering dental aesthetics, considering 'normative' or professional assessment alone (4). Increasingly, it is being recognized that patients' perceptions of need and outcome are important, if not essential, for measuring oral health and to this end a plethora of subjective oral health status measures have been developed (5).

The Oral Health Impact Profile (OHIP) is one of the most comprehensive subjective oral health status measures (6). It has a sound theoretical framework based on an oral health model of 'disease/condition–impairment–disability–handicap' with seven domains as described by Locker (7). It consists of 49 items which measure both the frequency and severity of oral problems on physical, social and psychological well-being. Since its development over a decade ago, OHIP has been used in a number of epidemiological studies (8, 9) as an outcome measure for numerous cohort studies and in clinical trials for different treatment interventions (10–12). As some research settings do not permit the use of the full battery of 49 questions

in the instrument, a shorter form of OHIP, the OHIP-14 was developed using regression modeling and factor analysis (13). Slade (13) suggested that it would be necessary to evaluate the performance of OHIP-14 using other research designs, such as longitudinal studies and experimental trials. Another short form of OHIP has also been developed to accommodate a specific oral health state, edentulousness, using the item impact reduction method: OHIP-EDENT (14). Therefore, it can be assumed that different short forms of OHIP could be developed that are more appropriate to specific oral health states, because OHIP, with its 49 items, can be considered as a generic oral health status measure rather than a condition-specific measure (15).

There are numerous methodological (statistical and expert-based) approaches for developing short-form measures of existing health status measures as apposed to developing new measures which is a laborious and expensive task (16). As it is not uncommon that different methods produce different short forms from the same dataset, it is important to compare the different versions and select the appropriate short form depending on its intended use, i.e. as a descriptive discriminatory instrument or an outcome tool (17).

The aim of this study was to develop and evaluate short forms of OHIP for dental aesthetics, employing both regression modeling and expert-based approaches. Moreover we wanted to compare the validity and reliability of the OHIP short forms developed in discriminating dental aesthetics, and to compare their sensitivities with respect to tooth whitening so as to propose a short-form version suitable for discriminating dental aesthetics and evaluating dental aesthetic outcomes, such as tooth whitening.

Methods

Subjects

A total number of 157 subjects who were dissatisfied with the colour of their teeth were invited to a clinical screening for this study. For inclusion, subjects on clinical examination had to have no dental caries, missing teeth, malalignment or dental restorations among their anterior teeth, periodontal pockets, or exogenous staining of their teeth. Eighty-seven subjects met the inclusion criteria and were provided with an array of tooth-whitening products available commercially (toothpaste, adhesive strip and paint-on gel containing various concentrations of hydrogen peroxide) to use at home according to manufacturer's instructions, and were invited to attend an evaluation examination 8 weeks later. Compliance was assessed by asking participants to maintain a diary of product use and to return the unfinished products at the end of the study period. The study protocol was approved by the Ethics Committee of the Faculty of Dentistry, The University of Hong Kong.

Data collection

Subjects self-completed a Chinese version of OHIP, which is similar in all respects to the original OHIP except for the obvious linguistic difference (18). For each of the 49 OHIP questions, subjects were asked to rate how frequently they had experienced an impact in the past '2 weeks' at the baseline (prior to receiving tooth-whitening products) and at evaluation 8 weeks later (on completion of use according to manufacturer's instructions). Responses were made on a Likert scale and coded 'very often' (score 4), 'fairly often' (score 3), 'occasionally' (score 2), 'hardly ever' (score 1) and 'never' (score 0), similar to the original OHIP. In addition, subjects self-rated two global questions relating to their overall dental appearance and their overall oral health status.

Development of short forms of OHIP for dental aesthetics

In order to develop a new short form of OHIP for aesthetics based on regression analysis, the following methods were employed on the baseline data. Principal component factor analysis was undertaken to identify a set of underlying factors contributing to OHIP responses. This was followed by promax rotation of the factors that accounted for the greatest amount of variation, and computation of factor loadings for each question to identify any that exceeded 0.4, which was used as a threshold for moderate to high loadings. For the factor analysis, communalities were set to the variable's squared multiple correlation with all other variables. Then, least-squares regression was used with the total OHIP score (obtained by summing the coded Likert responses from all questions) as the dependent variable and each question was used as an independent variable. A default stepwise procedure was used in which individual questions were considered sequentially for their contribution to R^2, and the first 14 items making the largest addition to R^2 were selected. A second procedure

imposed greater control over the stepwise procedure: again, items making the greatest contribution to total R^2 were added sequentially, except that no more than two items from each conceptual domain were permitted to enter the model. This controlled regression procedure was also conducted until 14 items, two from each of the seven domains, were selected.

The other method used to develop a short form of OHIP for dental aesthetics was an expert-based approach. The content of the original OHIP-49 was carefully analyzed to detect areas of redundancy or uselessness with respect to dental aesthetics, and then 14 questions, two questions from each domain, were selected from the OHIP-49, because the aspects that the questions explored were logically related to the aesthetics, for example, the question 'have you felt that your appearance has been affected because of problems with your teeth, mouth or dentures?' was selected.

Evaluating the short forms of OHIP developed for dental aesthetics

Summary scores were calculated for all subjects by summing all 49 OHIP statements (range 0–196), the subset of 14 items in the shortened version developed by Slade (7) (range 0–56), the subset of 14 items in the Chinese shortened version modified by Wong et al. (18) (range 0–56) which has different domain questions to Slade's OHIP-14 and only has five similar questions, the subset of 14 items in the shortened version developed by the regression analysis (range 0–56) and the subset of 14 items in the shortened version developed by the expert-based approach (range 0–56).

In determining the validity and reliability of the OHIP short forms developed (OHIP-regression and OHIP-conceptual) as discriminatory measures of dental aesthetics, their association with global ratings of overall dental appearance and overall oral health status was assessed, and compared with other OHIP forms (original OHIP-49, Slade's OHIP-14 and a Chinese short form of OHIP). In addition, the internal reliability of the OHIP short forms developed was assessed and compared with the other OHIP forms. In determining the responsiveness of the OHIP short forms developed, paired *t*-tests were performed to detect mean changes in OHIP scores with respect to the tooth-whitening intervention and these were compared with changes in other OHIP form scores. The sensitivity of the various OHIP measures was assessed by determining distribution changes in OHIP score; an indication of the magnitude of the statistical change was assessed by determinzing their effect size (ES), by dividing the mean observed change by the standard deviation of baseline score. Cohen (19) has described ES of 0.2 = small, <0.6 = moderate, and >0.8 = large.

Results

Short forms of OHIP developed

At baseline there were 87 subjects, and 17 of 49 OHIP questions had factor loadings that exceeded 0.4 for the first rotated factor. Factor loadings for the questions with the 17 highest loadings are presented in Table 1 (column headed 'Factor loading'). The high loading questions ranged in prevalence from 80.9% (Q7) to 0% (Q39, Q40, Q41, Q42) and severity from 2.08 (Q7) to 0.13 (Q42). The two disability domains and the handicap domain dominated this set of 17 questions. In the default stepwise regression procedure, the total R^2 was 0.97; no question was selected from the physical disability and the social disability domains. The prevalence of the 14 questions ranged from 47.6% (Q4) to 1.6% (Q48) and the severity scores ranged from 1.41 (Q4) to 0.21 (Q48) (Table 1). In the controlled regression procedure, in which entry into the model was limited to two questions from each domain, the prevalence of the 14 items ranged from 60.4% (Q3) to 0% (Q39) while the severity ranged from 1.71 (Q3) to 0.24 (Q39) (Table 1). The total R^2 for this set of questions was 0.90. This created the OHIP-regression short form.

On analyzing the content of the original OHIP-49 to detect areas of redundancy with respect to dental aesthetics, 14 questions, two from each domain were selected from the OHIP-49. Table 1 illustrates questions chosen to form the OHIP-conceptual short form.

Discriminatory ability of the various OHIP forms: validity and reliability

At baseline, 26% (23) rated their overall dental appearance as poor and 74% (64) as very poor. Rating of overall dental appearance was associated with OHIP-49 scores ($P = 0.026$), OHIP-conceptual scores ($P = 0.001$) and OHIP-regression scores ($P = 0.001$) (Table 2). With respect to global rating of oral health status, 15% (13) rated it poor, 53% (46) as OK, and 32% (28) as good. Global rating of oral health status was associated only with

Table 1. Prevalence, mean values, factor analysis and regression analysis for OHIP 49 (item level)

Conceptual dimension and item	Prevalence: % reporting item occasionally, fairly often or very often	Severity: item mean (0–4)	Factor loading	Sequential R^2 for: Default selection	Sequential R^2 for: Controlled selection
Functional limitation					
Q1 Difficulty chewing	19.10	0.95			
Q2 Trouble pronouncing word	9.50	0.48			
Q3 Noticed tooth that doesn't look right	60.40	1.71			−0.07
Q4 Appearance affected	47.60	1.41		0.10	0.08
Q5 Breath stale	36.50	1.27	0.42		
Q6 Taste worsens	3.20	0.52		0.10	
Q7 Food catching	80.90	2.08	0.43		
Q8 Digestion worse	14.30	0.73			
Q9 Dentures not fitting	3.20	0.16			
Physical pain					
Q10 Painful aching	19.10	0.92		0.15	
Q11 Sore jaw	12.70	0.89		0.07	
Q12 Headaches	9.50	0.62			
Q13 Sensitive teeth	28.60	1.10		0.10	0.14
Q14 Toothache	12.70	0.83		0.11	
Q15 Painful gums	22.20	1.02			0.24
Q16 Uncomfortable to eat	14.30	0.78			
Q17 Sore spots	14.30	0.84		0.16	
Q18 Discomfort (dentures)	0.00	0.10			
Psychological discomfort					
Q19 Worried	12.70	0.79			0.13
Q20 Self-conscious	14.30	0.84			
Q21 Miserable	6.30	0.56			
Q22 Appearance	50.80	1.48			0.11
Q23 Tense	11.10	0.78		0.13	
Physical disability					
Q24 Speech unclear	3.20	0.49			
Q25 Others misunderstood	3.20	0.30			
Q26 Less flavour in food	3.20	0.40			
Q27 Unable to brush teeth	12.70	0.75	0.44		
Q28 Avoid eating	12.70	0.56			0.12
Q29 Diet unsatisfactory	3.20	0.41			
Q30 Unable to eat (dentures)	0.00	0.10			
Q31 Avoid smiling	12.70	0.70	0.52		0.14
Q32 Interrupt meals	6.30	0.32	0.42		
Psychological disability					
Q33 Sleep interrupted	3.20	0.33			
Q34 Upset	7.90	0.57	0.45	0.16	0.09
Q35 Difficult to relax	3.20	0.43	0.65		
Q36 Depressed	3.20	0.32		0.12	
Q37 Concentration affected	4.80	0.41	0.44		
Q38 Been embarrassed	14.30	0.75	0.53	0.10	0.08
Social disability					
Q39 Avoid going out	0.00	0.24	0.68		0.12
Q40 Less tolerant of others	0.00	0.25	0.70		
Q41 Trouble getting on with others	0.00	0.17	0.63		
Q42 Irritable with others	0.00	0.13	0.66		
Q43 Difficulty doing jobs	4.80	0.32			0.23
Handicap					
Q44 Health worsened	4.80	0.41	0.45		0.06
Q45 Financial loss	25.40	0.81			0.09
Q46 Unable to enjoy people's company	3.20	0.49	0.61	0.14	
Q47 Life unsatisfying	6.30	0.46	0.46	0.10	
Q48 Unable to function	1.60	0.21	0.49	0.14	
Q49 Unable to work	3.20	0.21			

Table 2. Discriminatory ability of all the OHIP versions in assessing dental aesthetics

	Dental appearance			Oral health status			
	Very poor [mean (SD)]	Poor [mean (SD)]	*P*-value	Poor [mean (SD)]	OK [mean (SD)]	Good [mean (SD)]	*P*-value
OHIP-49	34.52 (17.91)	25.81 (14.08)	0.034	43.46 (17.41)	30.30 (17.63)	29.14 (16.06)	0.034
OHIP-14	6.26 (4.28)	7.87 (5.50)	0.175	10.23 (6.35)	6.78 (5.19)	7.04 (4.73)	0.108
OHIP-conceptual	12.87 (5.99)	7.48 (4.73)	<0.001	14.85 (5.52)	11.67 (6.31)	9.50 (5.99)	0.034
OHIP-regression	13.96 (6.68)	8.19 (4.95)	<0.001	16.15 (5.38)	12.67 (7.21)	10.14 (6.29)	0.029
OHIP-Chinese	7.65 (5.22)	6.67 (4.97)	0.401	9.46 (5.19)	6.70 (5.24)	7.32 (5.09)	0.242

OHIP-conceptual scores ($P = 0.035$), OHIP-regression scores ($P = 0.029$), and OHIP-49 scores ($P = 0.034$) (Table 2). In terms of internal reliability, Cronbach's alpha values of the OHIP-49 was 0.94, for Slade's OHIP-14 was 0.86, OHIP-conceptual was 0.86, OHIP-regression was 0.86, and OHIP-Chinese was 0.86.

Responsiveness of the various OHIP forms

Of the 87 subjects at baseline, 63 returned for the review appointment 8 weeks later. None presented with any signs or symptoms of adverse effects to general or oral health. Among them, one subject did not complete the questionnaire in full and was not considered in the analysis. This brought the overall response rate to 71% (62/87). The mean age of the group was 20.5 years (SD 1.5) and 71% (44) were women, and there was no significant difference in their age–gender profile compared with those who participated at baseline: mean age 21.1 (SD 2.1) and 69% (60) women ($P > 0.05$).

The mean baseline score, the mean follow-up OHIP score, and the observed effect are presented in Table 3. OHIP-conceptual (developed by an expert-based approach) had the largest mean observed change of 2.29 and the OHIP-14 developed by Slade (14) had the smallest mean observed change of 1.30. Significant differences between the baseline total OHIP score and the follow-up score were found in most OHIP forms except OHIP-regression. At domain levels, OHIP-conceptual had a significant difference between the baseline score and the follow-up score in three domains (functional limitation and psychological disability and handicap) but all the others had only two or less domains that had significant differences in scores between baseline and follow up. OHIP-49 had significant differences in the functional limitation and psychological discomfort domains, OHIP-14 by Slade had significant differences in the psychological discomfort and psychological disability domains, the Chinese-specific OHIP-14 had significant difference in the functional limitation and psychological discomfort domains, and the OHIP-regression only had one significant domain difference, functional limitation.

Although the mean changes in scores between baseline and follow up for most of the OHIP versions were relatively large, their ES were moderate (from +0.25 to +0.35) according to Cohen's criteria (19). At the domain level, the ES of the OHIP-49 ranged from −0.01 (social disability) to +0.51 (functional limitation), and for the OHIP-14 (Slade), it ranged from −0.021 (physical disability) to +0.36 (psychological discomfort). For the newly developed OHIP-conceptual, the ES range was from +0.12 (social disability) to +0.34 (handicap), and for the OHIP-regression the range was from +0.08 (physical disability) to +0.33 (functional limitation). For the Chinese-specific OHIP-14, the ES ranged from −0.08 (social disability) to +0.40 (functional limitation) (Table 3).

Table 4 illustrates the results for the ES of all the OHIP versions and their domains. The largest ES was found in the functional limitation domain of OHIP-49 (+0.51) and the smallest in the social disability of OHIP-49 (−0.01). From the table, OHIP-conceptual had the largest ES among the various OHIP versions with respect to the total score, physical pain, physical disability, psychological disability, social disability, and handicap domain scores.

Discussion

This study aimed to develop and evaluate a shortened version of the OHIP that could be used to discriminate dental aesthetics and to evaluate outcomes for aesthetic intervention in dentistry, such as tooth whitening. The original OHIP-49 is often not practical in a clinical setting because of its length and also because many questions are irrelevant to specific oral health states. Although

Table 3. Sensitivity of the all the OHIP versions to the effects of tooth whitening

	Baseline [mean (SD)]	Follow up [mean (SD)]	*P*-value (log)	Observed effect [mean (SD)]	Effect size
OHIP-49 score (0–196)	31.33 (17.83)	25.81 (17.13)	0.046*	5.52 (18.45)	+0.31
Functional limitation (0–36)	9.32 (3.69)	7.41 (3.69)	0.001*	1.90 (3.63)	+0.51
Physical pain (0–36)	7.08 (4.39)	5.94 (4.09)	0.213	1.14 (5.65)	+0.26
Psychological discomfort (0–20)	4.44 (2.82)	3.51 (2.84)	0.006*	0.94 (2.83)	+0.33
Physical disability (0–36)	4.00 (3.66)	3.79 (3.51)	0.652	0.21 (3.67)	+0.06
Psychological disability (0–24)	2.79 (2.88)	2.22 (2.72)	0.097	0.57 (3.07)	+0.20
Social disability (0–20)	1.11 (1.63)	1.13 (1.82)	0.890	−0.02 (2.23)	−0.01
Handicap (0–36)	2.59 (2.78)	1.81 (2.42)	0.064	0.78 (2.92)	+0.28
OHIP-14 (Slade) score (0–56)	7.33 (5.19)	6.03 (5.09)	0.038*	1.30 (5.38)	+0.25
Functional limitation (0–8)	1.00 (0.95)	0.86 (0.91)	0.239	0.14 (1.01)	+0.15
Physical pain (0–8)	1.70 (1.34)	1.40 (1.17)	0.196	0.30 (1.55)	+0.22
Psychological discomfort (0–8)	1.62 (1.13)	1.22 (1.26)	0.006*	0.40 (1.40)	+0.36
Physical disability (0–8)	0.73 (0.95)	0.75 (0.84)	0.746	−0.02 (1.09)	−0.021
Psychological disability (0–8)	1.17 (1.14)	0.84 (1.04)	0.026*	0.33 (1.15)	+0.29
Social disability (0–8)	0.44 (0.78)	0.43 (0.76)	0.926	0.02 (0.98)	+0.03
Handicap (0–8)	0.67 (0.90)	0.54 (0.86)	0.340	0.13 (1.10)	+0.15
OHIP conceptual score (0–56)	11.17 (6.51)	8.89 (5.77)	0.019*	2.29 (5.83)	+0.35
Functional limitation (0–8) Q3 Noticed tooth that doesn't look right Q4 Appearance affected	3.13 (1.77)	2.54 (1.70)	0.012*	0.59 (1.35)	+0.33
Physical pain (0–8) Q13 Sensitive teeth Q17 Sore spots	1.94 (1.18)	1.60 (1.30)	0.075	0.33 (1.66)	+0.28
Psychological discomfort (0–8) Q20 Self-conscious Q22 Appearance	2.32 (1.47)	1.95 (1.41)	0.058	0.37 (1.35)	+0.25
Physical disability (0–8) Q26 Less favour in food Q31 Avoid smiling	1.10 (1.15)	0.89 (0.97)	0.180	0.21 (1.10)	+0.18
Psychological disability (0–8) Q35 Difficult to relax Q38 Been embarrassed	1.17 (1.14)	0.84 (1.04)	0.026*	0.33 (1.15)	+0.29
Social disability (0–8) Q40 Less tolerant of others Q43 Difficulty doing jobs	0.57 (0.84)	0.59 (0.85)	0.493	0.10 (1.04)	+0.12
Handicap (0–8) Q46 Unable to enjoy people's company Q47 Life unsatisfying	0.95 (1.08)	0.59 (0.85)	0.028*	0.37 (1.20)	+0.34
OHIP regression score (0–56)	11.86 (7.18)	9.75 (6.21)	0.066	2.11 (6.18)	+0.29
Functional limitation (0–8) Q3 Noticed tooth that doesn't look right Q4 Appearance affected	3.13 (1.77)	2.54 (1.70)	0.012*	0.59 (1.35)	+0.33
Physical pain (0–8) Q13 Sensitive teeth Q15 Painful gums	2.11 (1.35)	1.86 (1.28)	0.315	0.25 (1.65)	+0.19
Psychological discomfort (0–8) Q19 Worried Q22 Appearance	2.27 (1.48)	1.87 (1.31)	0.059	0.40 (1.33)	+0.27
Physical disability (0–8) Q28 Avoid eating Q31 Avoid smiling	1.25 (1.32)	1.16 (1.25)	0.623	0.10 (1.28)	+0.09
Psychological disability (0–8) Q34 Upset Q38 Been embarrassed	1.32 (1.32)	1.02 (1.24)	0.059	0.30 (1.34)	+0.23
Social disability (0–8) Q39 Avoid going out Q43 Difficulty doing jobs	0.56 (0.82)	0.48 (0.80)	0.467	0.08 (0.97)	+0.10
Handicap (0–8) Q44 Health worried Q45 Financial loss	1.22 (1.42)	0.83 (0.98)	0.083	0.40 (1.40)	+0.28

Table 3. Continued

	Baseline [mean (SD)]	Follow up [mean (SD)]	*P*-value (log)	Observed effect [mean (SD)]	Effect size
OHIP-14 Chinese score (0–56)	7.25 (5.28)	5.84 (5.34)	0.019*	1.41 (5.94)	+0.27
Functional limitation (0–8) Q1 Difficulty chewing Q2 Trouble pronouncing words	1.43 (1.21)	0.95 (1.04)	0.009*	0.48 (1.42)	+0.40
Physical pain (0–8) Q16 Uncomfortable to eat Q17 Sore spots	1.62 (1.25)	1.33 (1.27)	0.163	0.29 (1.68)	+0.23
Psychological discomfort (0–8) Q19 Worried Q21 Miserable	1.35 (1.15)	1.08 (1.15)	0.041*	0.27 (1.23)	+0.23
Physical disability (0–8) Q26 Less favour in food Q32 Interrupt meals	0.71 (0.92)	0.60 (0.77)	0.480	0.11 (1.06)	+0.12
Psychological disability (0–8) Q34 Upset Q38 Been embarrassed	1.32 (1.32)	1.02 (1.24)	0.059	0.30 (1.34)	+0.23
Social disability (0–8) Q39 Avoid going out Q41 Trouble getting on with others	0.41 (0.66)	0.46 (0.80)	0.877	−0.05 (0.97)	−0.08
Handicap (0–8) Q48 Unable to function Q49 Unable to work	0.41 (0.82)	0.40 (0.75)	1.000	0.02 (0.89)	+0.02

*Paired *t*-test.

shorter versions of OHIP have been developed – the OHIP-14 developed by Slade (13) and a Chinese version of OHIP-14 by Wong et al. (18) – it was imperative to test their validity as discriminatory measures of dental aesthetics and to determine whether they were responsive to the effects of interventions aimed to improve dental aesthetics compared with short forms developed specifically for dental aesthetics through regression and expert-based approaches.

Short-form measures developed

Two shorter versions of OHIP were developed through different methods in this study. The OHIP-regression was developed following the method used by Slade (13) in which the OHIP-14 was developed, through factor analysis, stepwise and controlled stepwise regression. From the results of factor analysis, only one major component dominated while a large number of conceptually important items would be eliminated, including all the pain, discomfort and social disability items. Regression analysis resulted in selection of items that had a greater range in prevalence and severity compared with the range obtained from factor analysis. With default regression procedure, an R^2 value of 0.97 was obtained for five domains out of seven, in line with the conceptual model described by Locker (7). By controlling the process, it was possible to retain two items from each dimension with only a small reduction in R^2 to 0.90. For the default regression method, not all the conceptually important items were included such as items in the physical disability and social disability domains. Therefore, only the short-form OHIP that was developed by controlled stepwise regression was considered.

The second short form was developed as suggested by Coste et al. (20), they suggested that when an original health status measure cannot be considered as a gold standard, an expert-based approach to shortening possibly helped by statistical considerations is preferable to a statistical approach only. Results of this study supported this. Some difficulty was encountered when items needed to be chosen from domains especially when the domain itself seemed peripherally distinct from dental aesthetics. Therefore, items were chosen carefully based on the events that the subject were likely to encounter during the intervention, for example, when they applied the whitening gel on their teeth may become sensitive, as a result, the item exploring sensitive teeth was chosen. These items are also conceptually relevant for other dental aesthetic treatment such as veneer or crowns.

Table 4. Effect sizes of OHIP-49, OHIP-14, OHIP-conceptual, OHIP-regression and Chinese short-form OHIP

Rank	Total score	Functional limitation	Physical pain	Psychological discomfort	Physical disability	Psychological disability	Social disability	Handicap
1	OHIP-conceptual (+0.35)	OHIP-49 (+0.51)	OHIP-conceptual (+0.28)	OHIP-14 (+0.36)	OHIP-conceptual (+0.18)	OHIP-conceptual (+0.29)	OHIP-conceptual (+0.12)	OHIP-conceptual (+0.34)
2	OHIP-49 (+0.31)	OHIP-Chinese (+0.40)	OHIP-49 (+0.26)	OHIP-49 (+0.33)	OHIP-Chinese (+0.12)	OHIP-14 (+0.29)	OHIP-regression (+0.10)	OHIP-regression (+0.28)
3	OHIP-regression (+0.29)	OHIP-conceptual (+0.33)	OHIP-Chinese (+0.23)	OHIP-regression (+0.27)	OHIP-regression (+0.08)	OHIP-regression (+0.23)	OHIP-Chinese (−0.08)	OHIP-49 (+0.28)
4	OHIP-Chinese (+0.27)	OHIP-regression (+0.33)	OHIP-14 (+0.22)	OHIP-conceptual (+0.25)	OHIP-49 (+0.06)	OHIP-Chinese (+0.23)	OHIP-14 (+0.03)	OHIP-14 (+0.15)
5	OHIP-14 (+0.25)	OHIP-14 (+0.15)	OHIP-regression (+0.19)	OHIP-Chinese (+0.23)	OHIP-14 (−0.02)	OHIP-49 (+0.20)	OHIP-49 (−0.01)	OHIP-Chinese (+0.02)

Discriminatory ability of the various OHIP forms: validity and reliability issues

The short-form measures developed (OHIP-regression and OHIP-conceptual) were significantly associated with self perceptions of dental aesthetics and oral health status, unlike Slade's OHIP-14 and the Chinese short form of OHIP. Compared with OHIP-49, OHIP-regression and OHIP-conceptual had better discriminating ability with respect to self perceptions of dental aesthetics. These findings indicate the validity of the short-form measures developed as more appropriate for use in discriminating dental aesthetics. In terms of reliability, Cronbach's alpha values of the OHIP-regression and OHIP-conceptual short forms were similar (0.86) and comparable with Slade's OHIP-14 and the Chinese short form of OHIP and only marginally less than the OHIP-49 scale overall. Scales with Cronbach's alpha values above 0.70 are considered to have substantial reliability.

The prevalence of 'don't know' or 'blank' ranged from 0% to at most 6.3% which is low compared with what Slade had indicated when the OHIP-14 was developed (13). This indicates that the face validity of the newly developed instruments was high. Furthermore, the prevalence of reporting item 'occasionally', 'fairly often' or 'very often' ranged from 0% to 80.9% indicating that the oral health-related quality of life of the subjects was poor to start with, therefore there was little 'flooring effect' of this dataset, and thus there was room for improvement, especially for those questions which were conceptually related to aesthetics, such as 'noticed tooth that doesn't look right' (60.4%) and 'appearance' (50.8%).

Sensitivity of the different OHIP forms

Significant differences between the baseline total OHIP score and the follow-up score were found in most OHIP forms except OHIP-regression. At domain levels, OHIP-conceptual had a significant difference between the baseline score and the follow-up score in three domains (functional limitation and psychological disability and handicap) but among all the other OHIP forms significant differences were only observed among two or fewer domains. Significant differences in functional limitation and psychological discomfort was observed with OHIP-49, in psychological discomfort and psychological disability with Slade's OHIP-14, in functional limitation and psychological discomfort with the Chinese-specific OHIP-14,

and only in functional limitation with OHIP-regression.

When interpreting the magnitude of change based on the ES, the most popular approach has been to use Cohen's standardized ES (19), in which the mean change is divided by the standard deviation to serve as an 'effect size index'. The sensitivity of the different versions of OHIP was compared with respect to their ability as an outcome tool of dental aesthetics, in this case, tooth-whitening treatment. The largest effect, of the original OHIP was 0.51 in the functional limitation, which can just be considered as 'moderate' according to Cohen's criteria, and the effect size of the other OHIP forms could not be categorized as any larger. However, the subset derived by the expert-based approach (OHIP-conceptual) ranked highest in five of seven domains and also ranked highest when the ES of the total scores were compared. The OHIP-regression was only observed to have better sensitivity in the psychological discomfort domain, but the other domains demonstrated less sensitivity than the OHIP-conceptual. The relatively poor ES of the OHIP-14 and the Chinese short form of OHIP suggested that they are less appropriate for assessing dental aesthetics outcomes, as illustrated in this study on the effects of tooth whitening. The OHIP-conceptual appears more appropriate than OHIP-49 and other OHIP forms when it is used to detect changes in dental aesthetics, especially tooth colour. However, the sensitivity and responsiveness of OHIP-aesthetics (expert-based approach) should further be tested in clinical trials with other dental aesthetics treatments.

In conclusion, a short form of the OHIP derived in this study (OHIP-conceptual) was comparable with the full 49-item form in terms of measurement properties, was better than previous short forms developed for discriminatory dental aesthetics, and was more sensitive to measuring changes in dental aesthetics – tooth whitening.

References

1. Bruton PA, Ellwood R, Davies R. A six-month study of two self-applied tooth whitening products containing carbamide peroxide. Oper Dent 2004;29:623–6.
2. Garcia-Godoy F, Villalta P, Barjer ML, Gerlach RW. Placebo-controlled, 6-week clinical trial on the safety and efficacy of a low-gel, 14% hydrogen-peroxide whitening strip. Compend Contin Educ Dent 2004;25:21–6.
3. Claydon NC, Morman J, Bosma ML, Shirodaria S, Addy M, Newcombe R. Clinical study to compare the effectiveness of a test whitening toothpaste with a commercial whitening toothpaste at inhibiting dental stain. J Clin Periodontol 2004;31:1088–91.
4. Alkhatib MN, Holt R, Bedi R. Prevalence of self-assessed tooth discolouration in the United Kingdom. J Dent 2004;32:561–6.
5. Slade GD (ed). Measuring oral health related quality of life. Chapel Hill, NC: University of North Carolina; 1997.
6. Slade GD, Spencer AJ. Development and evaluation of the Oral Health Impact Profile. Community Dent Health 1994;11:3–11.
7. Locker D. Measuring oral health: a conceptual framework. Community Dent Health 1988;5:3–18.
8. Slade GD, Spencer AJ, Locker D, Hunt RJ, Strauss RP, Beck JD. Variations in the social impact of oral conditions among older adults in South Australia, Ontario and North Carolina. J Dent Res 1996;75:1439–50.
9. McGrath C. Oral health behind bars: a study of oral disease and its impact on the life quality of an older prison population. Gerodontology 2002;19:109–14.
10. 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.
11. John MT, Slade GD, Szentpetery 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;17:503–11.
12. McGrath C, Comfort MB, Lo EC, Luo Y. Can third molar surgery improve quality of life? A 6-month cohort study. J Oral Maxillofac Surg 2003;61:759–63.
13. Slade GD. Derivation and validation of a short-form oral health impact profile. Community Dent Oral Epidemiol 1997;25:284–90.
14. 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 Prosthodont 2002;15:446–50.
15. McGrath C, Bedi R. The value and use of 'quality of life' measures in the primary dental care setting. Prim Dent Care 1999;6:53–7.
16. Bowling A. Measuring health: a review of quality of life measurement scales Buckingham. Philadelphia, PA: Open University Press; 1997.
17. Locker D. Oral health and quality of life. Oral Health Prev Dent 2004;2:247–53.
18. Wong MC, Lo EC, McMillan AS. Validation of a Chinese version of the Oral Health Impact Profile (OHIP). Community Dent Oral Epidemiol 2002;30:423–30.
19. Cohen J. Statistical power analysis for the behavioural sciences, 2nd edn. Hillside, NJ: Lawrence Erlbaum and Associates; 1988.
20. Coste J, Guillemin F, Pouchout J, Fermanina J. Methodological approaches to shortening composite measurement scales. J Clin Epidemiol 1997;50:247–52.